PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classificati n 6 : C12N 15/12, C07K 14/47, 16/18, A01K 67/027, C12Q 1/68

A1

(11) Internati nal Publication Number: **WO 99/09169**

(43) Internati nal Publication Date: 25 February 1999 (25.02.99)

(21) International Application Number: PCT/US98/17255

(22) International Filing Date: 20 August 1998 (20.08.98)

(30) Priority Data:
60/056,217 21 August 1997 (21.08.97) US

(71) Applicant *(for all designated States except US)*: THE GOVERNMENT OF THE UNITED STATES OF AMERICA as represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: KASTNER, Daniel, L. [US/US]; 10112 Ashburton Lane, Bethesda, MD 20817 (US). AKSENTIJEVICH, Ivona [US/US]; 5719 Beech Avenue, Bethesda, MD 20817 (US). CENTOLA, Michael [US/US]; 915 Prospect Street, Tacoma Park, MD 20912 (US). DENG, Zuoming [CN/US]; Apartment 304, 18207 Lost Knife Circle, Gaithersburg, MD 20879 (US). SOOD, Raman [CA/US]; 10203 Nolan Drive, Rockville, MD 20850 (US). COLLINS, Francis, S. [US/US]; 5908 Tudor Lane, Rockville, MD 20852 (US). BLAKE, Trevor [US/US]; 19814 Falling Spring Court, Laytonsville, MD 20882 (US). LIU, P., Paul [US/US]; 7725 Blueberry Hill Lane, Ellicott City, MD 21043 (US). FISCHEL-GHODSIAN, Nathan [US/US]; Unit 104, 2122 Century Park Lane, Los Angeles, CA 90067 (US). GUMUCIO, Deborah, L. [US/US]; University of Michigan, Dept. of Anatomy and Cell Biology, 5793A Medical Sciences II, Ann Arbor, MI 48109-0616 (US). RICHARDS, Robert, I. [AU/AU]; 228 Brougham Place, North Adelaide, S.A. 5006 (AU). RICKE, Darrell, O. [US/US]; 4498A Fairway Drive, Los Alamos, NM 87544 (US). DOGGETT, Norman, A. [US/US]; P.O. Box 839, Santa Cruz, NM 87567 (US). PRAS, Mordechai [IL/IL]; Heller Institute of Medical Research, Chaim Sheba Medical Center, 52621 Tel-Hashomer (IL).

(74) Agent: BRUESS, Steven, C.; Merchant, Gould, Smith, Edell, Welter & Schmidt, P.A., 3100 Norwest Center, 90 South Seventh Street, Minneapolis, MN 55402-4131 (US).

(81) Designated States: AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: THE PYRIN GENE AND MUTANTS THEREOF, WHICH CAUSE FAMILIAL MEDITERRANEAN FEVER

(57) Abstract

The invention provides the nucleic acid sequence encoding the protein associated with familial Mediterranean fever (FMF). The cDNA sequence is designated as *MEFV*. The invention is also directed towards fragments of the DNA sequence, as well as the corresponding sequence for the RNA transcript and fragments thereof. Another aspect of the invention provides the amino acid sequence for a protein (pyrin) associated with FMF. The invention is directed towards both the full length amino acid sequence, fusion proteins containing the amino acid sequence and fragments thereof. The invention is also directed towards mutants of the nucleic acid and amino acid sequences associated with FMF. In particular, the invention discloses three missense mutations, clustered in within about 40 to 50 amino acids, in the highly conserved rfp (B30.2) domain at the C-terminal of the protein. These mutants include M680I, M694V, K695R, and V726A. Additionally, the invention includes methods for diagnosing a patient at risk for having FMF and kits therefor.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

THE PYRIN GENE AND MUTANTS THEREOF, WHICH CAUSE FAMILIAL MEDITERRANEAN FEVER

Background of the Invention

Field of the Invention

This invention relates to a novel genomic DNA sequence (*MEFV*) encoding a protein (pyrin) associated with familial Mediterranean fever (FMF). More specifically, the invention relates to the isolation and characterization of *MEFV*, and the correlation of mutations in *MEFV* with FMF disease.

Background of the Invention

Familial Mediterranean Fever (FMF) is a recessively inherited disorder characterized by dramatic episodes of fever, serosal inflammation and abdominal pain. This inflammatory disorder is episodic, with self-limited bouts of fever accompanied by unexplained arthritis, sterile peritonitis, pleurisy and/or skin rash. Patients often develop progressive systemic amyloidosis from the deposition of the acute phase reactant serum amyloid A (SAA). In some patients, progressive systemic amyloidosis can lead to kidney failure and death. The factors which incite an episode are unclear.

FMF is observed primarily in individuals of non-Ashkenazi Jewish, Armenian, Arab and Turkish background. Although rare in the United States, incidence of FMF in Middle Eastern populations can be as high as 1:7 in Armenian populations and 1:5 in non-Ashkenazi Jewish populations.

FMF attacks are characterized by a massive influx of polymorphonuclear leukocytes (PMNs) into the affected anatomic compartment. At the biochemical level, patients have been reported to have abnormal levels of C5a inhibitor (Matzner and Brzezinski, "C5a-inhibitor deficiency in peritoneal fluids from patients with familial Mediterranean fever," N. Engl. J. Med., 311:287-290 (1984)), neutrophil-stimulatory dihydroxy fatty acids (Aisen et al, "Circulating hydroxy fatty acids in familial Mediterranean fever," Proc. Natl. Acad. Sci. USA, 2:1232-1236 (1985)), and dopamine β-hydroxylase (Barakat et al, "Plasma dopamine beta-hyroxylase: rapid diagnostic test for recurrent hereditary polyserositis," Lancet, 2:1280-1283 (1988)). Although linkage studies have placed the gene causing FMF (designated

MEFV) on chromosome 16p (Pras et al., "Mapping of a gene causing familial Mediterranean fever to the short arm of chromosome 16," N. Engl. J. Med., 326:1509-1513 (1992); Shohat et al., "The gene for familial Mediterranean fever in both Armenians and non-Ashkenazi Jews is linked to the α-globin complex on 16p: evidence for locus homogeneity," Am. J. Hum. Genet., 51:1349-1354 (1992); Pras et al, "The gene causing familial Mediterranean fever maps to the short arm of chromosome 16 in Druze and Moslem Arab families," Hum. Genet., 94:576-577(1994); French FMF Consortium, "Localization of the familial Mediterranean fever gene (FMF) to a 250 kb-interval in non-Ashkenazi Jewish founder haplotypes," Am. J. Hum. Genet., 59:603-612(1996)), the genetic basis of FMF has not previously been identified.

Current treatment regimens for FMF include daily oral administration of colchicine. Although colchicine has been shown to cause near complete remission in about 75% of FMF patients and prevent amyloidosis, colchicine is not effective in all patients. Therefore, there is a need for new treatments for colchicine-resistant patients.

Additionally, there is a need for an accurate diagnostic test for FMF. Patients having FMF in countries where the disease is less prevalent often experience years of attacks and several exploratory surgeries before the correct diagnosis is made.

Summary of the Invention

The invention provides a novel genomic nucleic acid sequence (*MEFV*) **[SEQ ID NO: 1]**, shown in Figure 1, encoding the protein pyrin which is associated with familial Mediterranean fever (FMF). The corresponding cDNA sequence (v75-1) **[SEQ ID NO: 2]** and encoded amino acid sequence **[SEQ ID NO: 3]** are shown in Figure 2. The invention is also directed towards fragments of the DNA sequence that are useful, for example, as hybridization probes for diagnostic assays or oligonucleotides for PCR priming. Additionally, the invention is directed towards the corresponding sequence for the RNA transcript and fragments thereof.

Another aspect of the invention provides the amino acid sequence for a protein associated with FMF. This protein is called pyrin, to connote its relationship

to fever. The invention is directed towards both the full length amino acid sequence, fusion proteins containing the amino acid sequence and fragments thereof. These proteins are useful, for example, as antigens to produce specific anti-pyrin antibodies to be used as agents in diagnostic assays. Alternatively, the protein may be used in therapeutic compositions.

Mutations in pyrin result in FMF. Therefore, the invention is also directed towards mutants of the nucleic acid and amino acid sequences associated with FMF. In particular, the invention discloses three missense mutations, clustered in within about 40 to 50 amino acids, in the highly conserved rfp (B30.2) domain **[SEQ ID NO: 5]** at the C-terminal of the protein. These mutants include M680I, M694V, K695R and V726A, each of which is associated with FMF.

Additionally, the invention includes methods for diagnosing a patient at risk for having FMF using the nucleic acid and/or amino acid sequences of the invention. Such methods include, for example, hybridization techniques using nucleic acid sequences, PCR-amplification of *MEFV*, and immunoassays using anti-pyrin antibodies to identify mutations is *MEFV* or pyrin which are indicative of FMF.

Brief Description of the Figures

Figure 1 shows the genomic nucleic acid sequence for the gene associated with FMF;

Figure 2 shows a cDNA sequence and deduced amino acid sequence corresponding to the gene associated with FMF;

Figure 3 is a schematic representation of *MEFV* on chromosome 16p13.3;

Figure 4 show the expression profile of V75-1;

Figure 5 shows the DNA sequences of the M6801, M694V and V726A mutants; and

Figure 6 shows the alignment of multiple protein sequences with the C-terminal end of human pyrin.

Detailed Description of the Invention

The invention relates to the nucleic acid sequence encoding a protein associated with familial Mediterranean fever (FMF). The genomic DNA sequence is

designated *MEFV*. The corresponding cDNA sequence is designated as v75−1. The encoded protein is called pyrin, to connote its relationship to fever. The inventors have also discovered mutations in *MEFV* which are associated with FMF.

It is believed that pyrin is a nuclear factor that controls the inflammatory response in differentiated polymorphonuclear leukocytes (PMNs). In particular, pyrin is believed to be a negative autoregulatory molecule in PMNs. Knowledge of the genetic basis of FMF enables the production of diagnostic assays for FMF and treatments for FMF and other inflammatory diseases which are characterized by accumulation of PMNs, for example, acute infectious disease such as those caused by bacterial infection (e.g., *Pneumococcal* pneumonia), autoimmune diseases such as Sweets Syndrome or Behcet's disease, chronic arthritis, and the like.

The Nucleic Acid Sequence (*MEFV*)

The inventors have discovered the nucleic acid sequence for the gene associated with FMF. The nucleic acid sequence is found on chromosome 16p. Specifically, *MEFV* is located at 16p13.3 between the polycystic kidney disease gene (*PKD1*) and the tuberous sclerosis gene (*TSC2*) on the telomeric end, and the CREB−binding protein gene (*CREBBP*) on the centromeric end (see Figure 3).

The genomic DNA sequence encoding pyrin (*MEFV*) **[SEQ ID NO: 1]** is shown in Figure 1. The start methionine and stop codon are boxed, while the exons are underlined. The cDNA sequence (v75−1) **[SEQ ID NO: 2]** is shown in Figure 2. In Figure 2, the initial methionine and Kozak consensus sequences are underlined. The first boxed segment is a bZIP transcription factor basic domain. The second boxed segment is a Robbins/Dingwall consensus nuclear targeting signal. The segment indicated by +'s is a potential B−box zinc finger domain. The double−boxed region encloses a sequence which encodes a rfp, or B30.2, domain **[SEQ ID NO: 4]**. Within the double boxed region (the rfp or B30.2 domain), the nucleic acids encoding three FMF−associated mutations are double−underlined. Sites of synonymous single nucleotide polymorphisms are represented by the cents symbol "¢" above the sequence.

Although there is an excellent Kozak consensus sequence (Kozak, "Interpreting cDNA sequences: some insights from studies on translation," Mamm.

Genome, 7:563-574 (1996)) at the initial methionine (accATGG), the reading frame remains open in the cDNA upstream. Because there are no splice-acceptor consensus sequences or in-frame methionines with good Kozak sequences before the first stop upstream in the genomic DNA, the initial methionine remains the most likely starting methionine.

The RNA Transcript

The estimated transcript size from the nucleic acid sequence shown in Figure 2 is about 3503 nucleotides. The transcript size determined by Northern blotting is 3.7 kb. (See Example 4). The fact that the transcript size estimated from the sequence shown in Figure 2 approximates the size of the transcript found in experimental procedures further indicates that the sequence shown in Figure 2 is the full-length cDNA sequence.

The Encoded Protein

The inventors have also discovered the amino acid sequence for the protein associated with FMF (pyrin). Pyrin is predicted to be 781 amino acids in length and very positively charged. The pI is predicted to be greater than 8 (pI > 8), in part due to the fact that lysine and arginine residues make up 13% of the amino acid composition.

The predicted amino acid sequence for pyrin **[SEQ ID NO: 3]** is shown in Figure 2. The boxed segment from amino acid 266 to 280 is a bZIP transcription factor basic domain. The boxed segment from amino acid 420 to 437 is a Robbins/Dingwall consensus nuclear targeting signal. The segment indicated by +'s between residues 375 and 407 is a potential B-box zinc finger domain. The region double-boxed from residue 577 to 757 is a rfp, or B30.2, domain **[SEQ ID NO: 5]**. The rfp (B30.2) domain is conserved (sequence identity 40 - 60%) in molecules as diverse as butyrophilin (a milk protein with probable receptor function; Jack and Mather, "Cloning and molecular analysis of cDNA encoding bovine butyrophilin, an apical glycoprotein expressed in mammary tissue and secreted in association with the milk-fat globule membrane during lactation," J. Biol. Chem., 265:14481-14486 (1990)), A33 (a factor that binds polytene chromosomes in the newt; Bellini et al.,

"A putative zinc–binding protein on lampbrush chromosome loops," EMBO J., 12:107–114 (1993)), and xnf7 (a factor that binds mitotic chromosomes in the frog; Reddy et al., "The cloning and characterization of a maternally expressed novel zinc finger nuclear phosphoprotein (xnf7) in Xenopus laevis," Dev. Biol., 148:107–116 (1991)) and, by an analysis with the SEG algorithm (Wootton, "Non–globular domains in protein sequences: automated segmentation using complexity measures," Comput. Chem., 18:269–285 (1994)), most likely assumes a globular conformation. Within the double boxed region (the rfp or B30.2 domain), three of the amino acids that have been found mutated in FMF patients are double–underlined.

Positions of secondary structural elements were predicted by the profile neural network method PHDsec (Rost and Sander, "Prediction of protein secondary structure at better than 70% accuracy," J. Mol. Biol., 232:584–599 (1993); Rost and Sander, "Combining evolutionary information and neural networks to predict protein secondary structure," Proteins, 19:55–72 (1994)). The secondary structural elements in wild type pyrin (all β–sheets) as are shown as bold, horizontal arrows in Figure 6.

Expression

Pyrin is predominantly expressed in mature granulocytes and/or serosal cells. As shown in the Northern blots in Figure 4, high levels of pyrin are expressed in peripheral blood leukocytes (granulocytes), but not in lymph nodes, bone marrow, monocytes, lymphocytes, spleen or thymus (See Figure 4). Because granulocytes accumulate in tissues experiencing inflammation during a FMF episode, expression of pyrin in granulocytes is consistent with the clinical phenotype for FMF.

The restriction of pyrin to granulocytes, its apparent localization in the nucleus, and the phenotype associated with mutations tends to indicate that pyrin is a nuclear factor that controls the inflammatory response in differentiated PMNs. Additionally, the inventors found that pyrin shares homology with a number of molecules implicated in inflammation, such as rpt–1 (a known downregulator of inflammation). In view of the fact that FMF is a disease of excessive inflammation, and that pyrin shares homology to a known downregulator of inflammation, pyrin is believed to be a negative autoregulatory molecule in PMNs.

Homologies

Pyrin shares homology with a number of molecules implicated in inflammation including 52 kd Ro/SS A ribonucleoprotein (patients with systemic lupus erythematosus (SLE) and Sjögren's syndrome frequently make autoantibodies against this ribonucleoprotein); Staf-50 (an interferon-inducible transcriptional regulator; Tissot and Mechti, "Molecular cloning of a new interferon-induced factor that represses human immunodeficiency virus type 1 long terminal repeat expression," J. Biol. Chem., 270:14891-14898 (1995)); and rpt-1 (a mouse downregulator of IL-2; Patarca et al., "rpt-1, an intracellular protein from helper/inducer T cells that regulates gene expression of interleukin 2 receptor and human immunodeficiency virus type 1," Proc. Natl. Acad. Sci. USA, 85:2733-2737 (1988)).

The homology between pyrin and rpt-1 is found in a domain extending from residues 385 - 550 on pyrin. Pyrin shows particularly high homology to many proteins, including 50 kdRo/SS A and Staf-50, at the C-terminal end, the rfp (B30.2) domain. Figure 6 shows the alignment of the C-terminal end of human pyrin with multiple sequences having statistical similarity as assessed by BLAST (Altschul et al., *supra*). Search cutoffs used to identify homologs were a Karlin-Altschul score of two aligned sequences ≥ 70 with a probability $\leq 10^{-3}$. At each position, residues occurring in a majority of the sequences are shown in inverse type. The numbering scheme at the top of the figure is based on the sequence of pyrin.

The B-box zinc finger and rfp (B30.2) domain combination observed in pyrin is also seen in 52 kd Ro/SS A and *ret* finger protein. The spacing between the B-box zinc finger and the rfp (B30.2) domain is highly conserved, suggesting that precise orientation of the two domains with respect to one another may be required for function.

Mutants

The inventors have also discovered missense mutations that are found in individuals affected with FMF, but not found in any of a large panel of normal control chromosomes. The missense mutations are clustered within about 40 to 50

amino acids (including residues 680 through 726) in the highly conserved rfp (B30.2) globular domain. It is believed that the mutations affect the secondary structure of this domain and result in a structural change that prevents the normal pyrin-mediated negative feedback loop.

A first mutation associated with FMF is a G ⌫ C transversion at nucleotide 2040 which results in the substitution of isoleucine for methionine (M680I). A second mutation is an A ⌫ G transition at nucleotide 2080 which results in the substitution of valine for methionine (M694V). A third mutation is a T ⌫ C transition at nucleotide 2177 which results in the substitution of alanine for valine (V726A). Additionally, the inventors have discovered a fourth mutation at position 695 which results in the substitution of Arginine for Lysine (K695R).

It is believed that phenotypic variation in FMF may be attributable to the differences between mutations. For example, the M694V mutation is very common in populations with the highest incidence of systemic amyloidosis (especially North African Jews). On the other hand, V726A is seen in populations in which amyloid is less common (Iraqi and Ashkenazi Jews, Druze and Armenians).

Figure 5 shows DNA sequence electropherograms, produced by amplifying exon 10 genomic DNA and sequencing, which demonstrate the M680I, M694V, and V726A substitutions. For each mutation, individuals who are homozygous for the normal allele are shown at the top, heterozygotes between the normal and mutant allele are shown in the middle, and homozygotes for the mutation are shown at the bottom.

None of these mutations result in a truncated protein. This is consistent with the periodic nature of the inflammatory attacks in FMF. Other diseases with periodic episodes are associated with a protein that functions adequately at steady state, but decompensates under stress, such as sickle cell anemia (Weatherall et al., "The hemoglobinopathies," In The Metabolic and Molecular Bases of Inherited Disease, Scriver et al, eds., New York, McGraw-Hill, pp. 3417-3484 (1995) and hyperkalemic periodic paralysis (Ptacek et al., "Identification of a mutation in the gene causing hyperkalemic periodic paralysis," Cell, 67:1021-1027 (1991)).

Diagnostic Methods

The sequences provided by this invention can be used in methods for diagnosis of risk for developing FMF. As used herein, an individual is "at risk" for developing FMF when the individual has a mutant *MEFV* nucleic acid sequence which results in expression of mutant pyrin, particularly where the amino acid mutation occurs in the highly conserved rfp (B30.2) C-terminal domain. Mutations include substitutions of one nucleic acid with a different nucleic acid. In contrast, a patient having wild type *MEFV* nucleic acid sequence expressing wild type pyrin is not at risk for developing FMF. As used herein, "wild type" refers to a dominant genotype which naturally occurs in the normal population (i.e., members of the population not afflicted with familial Mediterranean fever). Thus, methods for identifying an individual's specific nucleic acid or amino acid sequence are useful for determining risk of FMF. Specifically, a method for determining whether an individual's nucleic acid sequence encodes a wild type or mutant pyrin is useful in determining whether the individual is at risk for developing FMF.

Many methods for analysis of an individuals nucleic acid or amino acid sequences are known to those of skill in the art, and include, for example, direct sequencing, ARMS (amplification refractory mutation system), restriction endonuclease assays, oligonucleotide hybridization techniques, and immunoassays. While some commonly used procedures are exemplified below, the inventors are aware that other methods are available and include them within the scope of their invention.

Southern Blot Techniques

In Southern blot analysis, DNA is obtained from an individual and then separated by gel electrophoresis. Following electrophoresis, the double stranded DNA is converted to single stranded DNA, for example, by soaking the gel in NaOH. The DNA is then transferred to a sheet of nitrocellulose. The DNA is then contacted with a labeled probe. For example, labeled probe can be applied to the nitrocellulose after it dries. As used herein, a "probe" is a nucleic acid sequence that is complementary to the sequence of interest. The probe can be either a DNA sequence or an RNA sequence. Preferably the probe is about 8 to 16 nucleotides in

length. A radioactive label, such as ^{32}P is an example of a suitable label. Other suitable labels include fluorophores or an enzyme which catalyzes a color producing reaction (e.g., horse radish peroxidase). Because the probe has complementary sequence to the DNA sequence of interest, it will hybridize to the specific DNA sequence. As used herein, "hybridize" means that the probe will form a double-stranded molecule with the specific DNA sequence by complementary base pairing under conditions of high stringency (e.g., 65°C; 0.1 x SSC; Sambrook et al., Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, New York: Cold Spring Harbor Press (1989)). After the probe is allowed to hybridize to the DNA, excess probe is washed away. The hybridized DNA is easily visualized from the labeled probe using known techniques. Hybridization of the probe indicates that the sample DNA contains a sequence that is complementary to the labeled probe. In a preferred method, hybridization probes are designed from the *MEFV* nucleic acid sequences, and particularly, from the C-terminal *MEFV* sequence encoding the rfp (B30.2) globular domain.

It is often desirable to amplify the sample DNA for more efficient analysis. Polymerase chain reaction (PCR) can be used to amplify the DNA. PCR is a technique that is well known to one of skill in the art. An exemplary method includes developing oligonucleotide primers that hybridize to opposite strands of DNA flanking the *MEFV* gene. As used herein, a "primer" is a short nucleotide sequence which is complementary to a DNA sequence flanking the DNA sequence of interest. Preferably the primer is about 15 to 20 nucleotides in length. The specific fragment defined by the primers exponentially accumulates by repeated cycles of denaturation, oligonucleotide primer annealing and primer extension. In a preferred embodiment, the PCR primers amplify the region encoding the rfp (B30.2) globular domain. The amplified domain can then be analyzed by hybridization or screening techniques.

For example, oligonucleotide primers are developed to amplify *MEFV*, the rfp (B30.2) domain, or a fragment thereof, such as the preferred 40 to 50 amino acid fragment of the rfp (B30.2) domain discussed above. Suitable oligonucleotide primers, such as "Exon 10A Forward and Reverse", "Exon 10B Forward and

Reverse", and "Exon 10B Forward and Exon 10A Reverse", are shown in Example 1.

Northern Blot Techniques

The presence of a wild type or mutant RNA transcript may be determined by Northern Blot Techniques, following a procedure similar to that outlined for the Southern Blot Technique.

Western Blot Techniques

The presence of a wild type or mutant protein from the highly conserved C-terminal rfp (B30.2) region can be detected by immunoassay, for example by Western Blot Techniques. In this procedure, a tissue sample is obtained from an individual and separated by gel electrophoresis. Following electrophoresis, the proteins are then transferred to nitrocellulose. The proteins are then contacted with a labeled probe, for example, by applying the labeled probe to the nitrocellulose after it is dried. Suitable probes include labeled anti-pyrin antibodies, preferably those antibodies specific for an epitope in the highly conserved C-terminal rfp (B30.2) domain. Exemplary labels include radioactive isotopes, enzymes, fluorophores and chromophores. Because it is believed that mutants in the highly conserved C-terminal domain alter the secondary structure of the domain, an antibody specific for the wild-type protein should not bind to or recognize a protein having a mutation in this highly conserved region. Conversely, an antibody specific for a mutant protein does not recognize or bind to the wild type. After excess antibody is rinsed away, the presence of the specific protein/antibody complex is easily determined by known methods, for example by development of the label attached to the anti-pyrin antibody, or by the use of secondary antibodies.

Sequencing Techniques

Alternately, DNA, RNA or protein obtained from an individual can be sequenced by known methods, and compared to the wild type sequence. Mutations recognized in the sequence, particularly, in the rfp (B30.2) domain indicate risk for developing FMF.

ARMS

ARMS (amplification refractory mutation system) is a PCR based technique in which an oligonucleotide primer that is complementary to either a normal allele or mutant allele is used to amplify a DNA sample. In one variation of this method, a pair of primers is used in which one primer is complementary to a known mutant sequence. If the DNA sample is amplified, the presence of the mutant sequence is confirmed. Lack of amplification indicates that the mutant sequence is not present. In a different variation, the primers are complementary to wild type sequences. Amplification of the DNA sample, indicated that the DNA has the wild type sequence complementary to the primers. If no amplification occurs, the DNA likely contains a mutation at the sequence where hybridization should have occurred. A description of ARMS can be found in Current Protocols in Human Genetics, Chapter 9.8, John Wiley & Sons, ed by Dracopoli et al. (1995).

Restriction Endonuclease Assays

Restriction endonuclease assays can also be used to screen a DNA sample for mutants, such assays are used by Pras et al., "Mutations in the SLC3A1 transporter gene in *Cystinuria*," Am. J. Hum. Genet., 56:1297-1303 (1995). Briefly, a DNA sample is amplified and then exposed to restriction endonucleases that will or will not cleave the DNA depending on whether or not a mutation is present. After cleavage, the size of restriction fragments are observed to determine whether or not cleavage occurred.

Oligonucleotide Hybridization Techniques

Hybridization techniques, such as dot blots, are known to one of skill in the art and can be used to determine whether a DNA sample contains a specific sequence. In a dot blot, a DNA sample is denatured and exposed to a labeled probe which is complementary for a wild type sequence or a mutant sequence. Hybridization of a probe that is complementary to the wild type sequence (a "wild type probe") indicates that the wild type sequence is present. If the wild type probe does not hybridize to the DNA in the sample, the wild typ sequence is not present.

In a variation of this technique a probe that is complementary to a know mutant sequence can be used. A discussion of allele specific oligonucleotide testing can be found in Current Protocols in Human Genetics, Chapter 9.4, *supra*.

Immunological Assays

An immunological assay, such as an Enzyme Linked Immunoassay (ELISA), can be used as a diagnostic tool to determine whether or not an individual is at risk for developing FMF. One of skill in the art is familiar with the procedure for performing an ELISA. Briefly, antibodies are generated against native or mutant pyrin. This can be accomplished by administering a native or mutant protein to an animal, such as a rabbit. The anti-pyrin antibodies are purified and screened to determine specificity. In one representative example of an immunoassay, wells of a microtiter plate are coated with the specific anti-pyrin antibodies. An aliquot of a sample from a patient to be analyzed for pyrin is added in serial dilution to each antibody coated well. The sample is then contacted with labeled anti-pyrin antibodies. For example, labeled anti-pyrin antibodies, such as biotinylated anti-pyrin antibodies, can be added to the microtiter plate as secondary antibodies. Detection of the label is correlated with the specific pyrin antigen assayed. Other examples of suitable secondary antibody labels include radioactive isotopes, enzymes, fluorophores or chromophores. The presence of bound labeled (biotinylated) antibody is determined by the interaction of the biotin with avidin coupled to peroxidase. The activity of the bound peroxidase is easily determined by known methods.

Production of Pyrin

The nucleic acid sequence encoding wild type or mutant pyrin can be used to produce pyrin in cells transformed with the sequence. For example, cells can be transformed by known techniques with an expression vector containing v75-1 cDNA sequence operably linked to a functional promoter. Expression of pyrin in transformed cells is useful *in vitro* to produce large amounts of the protein. Expression *in vivo* is useful to provide the protein to pyrin-deficient cells. Examples of suitable host cells include animal cells such as bacterial or yeast cells, for

example, E. coli. Additionally, mammalian cells, such as Chinese hamster ovary (CHO) cells can be used. Human cells, such as SW480 colorectal adenocarcinoma can also be used as host cells.

Due to degeneracy of the genetic code, most amino acids are encoded by more than one codon. Therefore, applicants recognize, and include within the scope of the invention, variations of the sequence shown in **SEQ ID NO: 1**. For example, codons in a DNA sequence encoding pyrin can be modified to reflect the optimal codon frequencies observed in a specific host. Rare codons having a frequency of less than about 20% in known sequences of the desired host are preferably replaced with higher frequency codons.

Additional sequence modifications are known to enhance protein expression in a cellular host. These include elimination of sequences including spurious polyadenylation signals, exon/intron splice site signals, transposon-like repeats, and other well characterized sequences which may be deleterious to gene expression. The G-C content of a sequence may be adjusted to levels average for a given cellular host, as calculated by reference to known genes expressed in the host cell. Where possible, the sequence is modified to avoid predicted hairpin secondary mRNA structures. The genomic sequence might additionally be modified by the removal of introns.

Transgenic Animals

The nucleic acid sequences encoding pyrin, both wild-type and mutant, provided in this application are useful for the development of transgenic animals expressing pyrin. Such transgenic animals are used, for example, to screen compounds for treating FMF or inflammation.

Useful variations of a transgenic animal are "knock out" or "knock in" animals. In a "knock out" animal, a known gene sequence, such as the sequence encoding pyrin, is deleted from the animal's genome. Experiments can be performed on the animal to determine what effect the absence of the gene has on the animal. In a "knock in" experiment, the wild type gene is deleted and a mutant version or a gene from another organism is inserted therefore. Experiments can be performed on the animal to determine the effects of this transition.

Kits

The invention is also directed towards a kit for diagnosing risk of FMF. A suitable diagnostic kit includes a nucleic acid sequence encoding wild-type pyrin and at least one nucleic acid sequence encoding mutant pyrin. An alternative kit includes an anti-pyrin antibody which binds to wild-type pyrin and at least one anti-pyrin antibody which binds to mutant pyrin. A kit also preferably contains at least one pair of amplification primers capable of amplifying a nucleic acid sequence encoding pyrin. Preferably, the primers amplify a nucleic acid sequence encoding a rfp (B30.2) domain of pyrin.

The present invention may be better understood with reference to the following examples. These examples are intended to be representative of specific embodiments of the invention, and are not intended as limiting the scope of the invention.

Examples

The DNA samples used in the following examples were extracted from whole blood or from Epstein-Barr virus-transformed lymphocytes by standard techniques. The DNA was obtained from forty-four families of non-Ashkenazi Jewish descent (18 Moroccan, 14 Libyan, 5 Tunisian, 2 Egyptian and 5 Iraqi) and 5 Arab/Druze families (identified and sampled at the Chaim Sheba Medical Center in Tel-Hashomer, Israel). Additionally, twelve Armenian families were recruited from Cedars-Sinai Medical Center in Los Angeles. One Ashkenazi/Iraqi Jewish family was also studied.

The diagnosis of FMF in all families was according to established clinical criteria (Sohar et al., "Familial Mediterranean fever: a survey of 470 cases and review of the literature," Am. J. Med., 43:227-253 (1967)).

Example 1. Positional Cloning

A positional cloning approach was used to clone a new cDNA (v75-1) from the FMF candidate region on chromosome 16p13.3. Mutational analysis indicates the v75-1 is the gene (designated *MEFV*) expressing pyrin, mutations of which are associated with FMF disorder.

Publicly available polymorphic markers (discussed below) were used to narrow the candidate region on chromosome 16p to an approximately 1 Mb interval between *D16S94* and *D16S2622* (Sood et al., "Construction of a 1-Mb restriction mapped cosmid contig containing the candidate region for the familial Mediterranean fever locus (*MEFV*) on chromosome 16p13.3," Genomics, 42:83-95 (1997)) lying between the polycystic kidney disease (*PKD1*) and tuberous sclerosis (*TSC2*) genes on the telomeric end, and the CREB-binding protein (*CREBBP*) gene on the centromeric end (see Figure 3). Because physical maps constructed around these genes did not extend into the *MEFV* region, a contig was constructed which spanned the candidate region.

Attempts to construct a mega YAC (yeast artificial chromosome) contig spanning the *MEFV* candidate region were unsuccessful due to the instability of YAC clones from this region of chromosome 16. Instead, a cosmid map was assembled by iterative screening of a flow sited chromosome 16 specific cosmid library. *D16S246* was the telomeric starting point of the chromosomal walk. Identification of recombinants at *D16S2622* enabled us to use this microsatellite marker as the centromeric boundary (Sood et al., 1997, *supra*).

Observed recombinations of microsatellite markers in a panel of 61 families defined a critical region of 285 kb (*D16S468 - D16S3376*).

By analysis of the genomic sequence from this region, two new microsatellites, *D16S3404* and *D16S3405* (Figure 3B), were found in the center of the *D16S3082 - D16S3373* interval. In one non-Ashkenazi Jewish family, evidence of a historical recombination event between *D16S3404* and *D16S3405* in the highly conserved non-Ashkenazi Jewish haplotype (designated haplotype A) was observed. Therefore, the region telomeric of *D16S3405* (and 4 candidate genes encoded

therein) were excluded from further consideration. The discovery of the two new microsatellites and the historical recombination event further refined of the candidate interval to the centromeric-most 115 kb.

A combined strategy of exon amplification, direct cDNA selection, and single-pass sequencing led to the isolation of 9 full length cDNA clones. The furthest centromeric cDNA clone, v75-1, was isolated by solution hybridization of a leukocyte cDNA library with biotinylated oligonucleotide probes derived from two exons trapped from PAC 273L24.

Exon Trapping

PAC (P1 artificial chromosome) clone 273L24 (Genome Systems; St. Louis) includes the centromeric-most 115 kb. Therefore, exon trapping was performed on PAC clone 273L24. Exon trapping was performed essentially as described by Buckler et al., "Exon amplification: a strategy to isolate mammalian genes based on RNA splicing," Proc. Natl. Acad. Sci. USA, 88:4005-4009 (1991). Essentially, PAC clone 273L24 was partially digested with Sau 3AI (commercially available, for example, from New England Biolabs). The reaction products were size fractionated by agarose gel electrophoresis and DNA fragments 2 kb and larger were isolated from the gel. Fifty ng of partially digested DNA was ligated with 10 ng of exon trapping vector pSPL3 (Exon Trapping System; Life Technologies, Gaithersburg, MD) that had been previously cleaved with Bam HI (commercially available) and dephosphorylated with calf intestinal alkaline phosphatase (Promega, Madison, WI). Ligation products were electroporated into E. coli DH12B (Life Technologies, Gaithersburg, MD) The electroporated cells were cultured en mass in LB broth with 200 mg/ml ampicillin for 16 hours at 37° C with shaking.

DNA prepared from the culture was used to transfect COS-7 cells (ATCC 30-2002) using lipofectACE reagent (Life Technologies, Gaithersburg, MD). Total RNA was isolated from transfected COS-7 cells with Trizol reagent (Life Technologies) followed by ethanol precipitation.

First strand cDNAs of transcription products from pSPL3 were primed with the oligonucleotide SA2 (Exon Trapping System; Life Technologies, Gaithersburg, MD). Specific amplification of trapped exons was as follows: PCR primed with

oligonucleotides SA2 and SD6 (Exon Trapping System; Life Technologies, Gaithersburg, MD) was performed, followed by digestion of the PCR products with Bst XI (commercially available).

A second PCR reaction using the digestion products was primed with oligonucleotides dUSD2 and dUSA4 (Exon Trapping System; Life Technologies, Gaithersburg, MD). The resulting DNA fragments were cloned into pAMP10 vector (Exon Trapping System; Life Technologies, Gaithersburg, MD) and sequenced. Two hundred clones were sequenced and 20 independent exons were identified by visual inspection and hybridization to DNA fragments from the FMF critical region, with several exons identified more than one time.

Oligonucleotides for Exon Amplification

Oligonucleotides used to amplify pyrin exons were as follows (all oligo sequences are given 5' to 3'):

Exon 1 forward, AAC CTG CCT TTT CTT GCT CA; **[SEQ ID NO: 6]**
Exon 1 reverse, CAC TCA GCA CTG GAT GAG GA; **[SEQ ID NO:7]**
Exon 2A forward, ATC ATT TTG CAT CTG GTT GTC CTT CC; **[SEQ ID NO:8]**
Exon 2A reverse, TCC CCT GTA GAA ATG GTG ACC TCA AG; **[SEQ ID NO:9]**
Exon 2B forward, GGC CGG GAG GGG GCT GTC GAG GAA GC; **[SEQ ID NO:10]**
Exon 2B reverse, TCG TGC CCG GCC AGC CAT TCT TTC TC; **[SEQ ID NO:11]**
Exon 3 forward, TGA GAA CTC GCA CAT CTC AGG C; **[SEQ ID NO: 12]**
Exon 3 reverse, AAG GCC CAG TGT GTC CAA GTG C; **[SEQ ID NO: 13]**
Exon 4 forward, TTG GCA CCA GCT AAA GAT GGC; **[SEQ ID NO: 14]**
Exon 4 reverse, TCT CCC TCT ACA GGG ATG AGC; **[SEQ ID NO: 15]**
Exon 5 forward, TAT CGC CTC CTG CTC TGG AAT C; **[SEQ ID NO: 16]**
Exon 5 reverse, CAC TGT GGG TCA CCA AGA CCA AG; **[SEQ ID NO: 17]**
Exon 6 forward, TCC AGG AGC CCA GAA GTA GAG; **[SEQ ID NO: 18]**
Exon 6 reverse, TTC TCC CTA TCA AAT CCA GAG; **[SEQ ID NO: 19]**

Exon 7 forward, AGA ATG TAG TTC ATT TCC AGC; **[SEQ ID NO: 20]**
Exon 7 reverse, CAT TTC TGA ACG CAG GGT TT; **[SEQ ID NO: 21]**
Exon 8/9 forward, ACC TAA CTC CAG CTT CTC TCT GC; **[SEQ ID NO: 22]**

Exon 8/9 reverse, AGT TCT TCT GGA ACG TGG TAG; **[SEQ ID NO: 23]**
Exon 10A forward, CCA GAA GAA CTA CCC TGT CCC; **[SEQ ID NO: 24]**
Exon 10A reverse, AGA GCA GCT GGC GAA TGT AT; **[SEQ ID NO: 25]**
Exon 10B forward, GAG GTG GAG GTT GGA GAC AA; **[SEQ ID NO: 26]**
Exon 10B reverse, TCC TCC TCT GAA ATC CAT GG. **[SEQ ID NO: 27]**

Direct cDNA selection

Direct cDNA selection was used to isolate 2 full-length cDNA clones (Parimoo et al., "cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments," Proc. Natl. Acad. Sci. USA, 88:9623-9627 (1991). Cosmids, BAC (bacterial artificial chromosome) and P1 clones in the FMF candidate region were biotinylated using BioPrime (Life Technologies, Gaithersburg, MD). cDNAs were prepared from combined mRNA from fetal brain, fetal liver, and human lymph node by reverse transcription and ligation of an EcoRI/NotI adaptor to second strand cDNAs.

cDNAs were directly hybridized to biotinylated templates which were recovered using streptavidin-labeled magnetic beads. Conditions for blocking, hybridization, binding and elution of cDNAs from magnetic beads (Dynal) were as described by Parimoo et al., *supra*. After two rounds of selection, eluted cDNAs were amplified with CUA-tailed EcoRI/NotI adaptor primers and subcloned into the pAMP10 vector (Life Technologies, Gaithersburg, MD) to yield libraries of selected cDNAs.

Recombinant clones were arrayed on blots. Clones that hybridized to either repetitive or ribosomal sequences were excluded from further analysis. To confirm their origin, unique clones were individually hybridized to EcoRI digests of cosmid/BAC/P1 DNAs and DNAs from chromosome 16-specific human-hamster hybrid lines. Clones were then hybridized to each other and were binned into groups. Representative clones of each group were hybridized to multiple tissue Northern blots and sequenced.

cDNA Identification by Solution Hybridization

Following the protocol provided in the Gene Trapper kit, the furthest centromeric cDNA, clone v75-1, was isolated by solution hybridization of a leukocyte cDNA library with biotinylated oligonucleotide probes derived from 2 exons trapped from PAC 273L24. Solution hybridization was carried out using the

GeneTrapper cDNA Positive Selection System (Life Technologies, Gaithersburg, MD).

Two trapped exons, v66 and v75, were used as starting material. PCR screening of Superscript cDNA libraries (Life Technologies, Gaithersburg, MD) derived from human brain, liver, leukocytes, spleen, and testis were used to determine the tissue-specific expression of these exons. GeneTrapper experiments were performed with sense and antisense primers from both exons, assuming both orientations of these exons in the putative transcript.

The following oligonucleotides were synthesized and PAGE-purified:

v66GT1: AAG CTC ACT GCC TTC TCC TC; **[SEQ ID NO: 28]**

v66GT2: GAG GAG AAG GCA GTG AGC TT; **[SEQ ID NO: 29]**

v75GT1: GAC TTG GAA ACA AGT GGG AG; **[SEQ ID NO: 30]**

v75GT2: CTC CCA CTT GTT TCC AAG TC. **[SEQ ID NO: 31]**

Oligos were biotinylated, hybridized to single-stranded DNA from the leukocyte cDNA library (one primer per reaction), followed by cDNA capture using paramagnetic streptavidin beads and repair using the corresponding non-biotinylated oligos. Colony hybridization of lifts using ^{32}P-dCTP end-labeled oligos was used to identify positive clones. Gel-purified inserts from these clones were hybridized to cosmid contig blots in order to distinguish cDNA clones mapping to the FMF region from false positive clones due to homologous domains. All positive clones were identified by the primers v66GT2 and v75GT2, and no clones were identified by the other set of primers.

Characterization of cDNA v75-1

The translated v75-1 cDNA sequence is shown in Figure 2. The exon-intron structure deduced from the genomic sequence of two cosmids is depicted in Figure

3C. Shaded boxes represent exons; introns are drawn to scale. The numbers above the boxes represent the size of the exons in bp. The numbers below the boxes reflect the order of the exons with 1 being the most 5'.

Although there is an excellent Kozak consensus (Kozak, *supra*) at the initial methionine, the reading frame remains open in the cDNA upstream. There are no splice-acceptor consensus sequences or in-frame methionines with good Kozak sequences before the first stop upstream in the genomic DNA. Additionally, the transcript size by Northern blot is 3.7 kb. The estimated transcript size from cDNA is 3503 nucleotides. Therefore, the sequence appears to be the full-length sequence.

Example 2. mutational analysis

Three different v75-1 mutants of FMF carrier chromosomes in multiple ethnic groups are not seen in a panel of almost 300 normal control chromosomes. This indicates that v75-1 is a cDNA of *MEFV*, the gene associated with FMF.

Three missense mutations were identified in exon 10 of v75-1 (Figure 5) after screening a total of 165 individuals from 65 families. All three mutations are clustered within 46 amino acids of one another in the highly conserved rfp (B30.2) globular domain at the C-terminal end of the predicted protein. The first mutation, is a G ⌫ C transversion at nucleotide 2040 in which methionine is replaced by isoleucine (M680I). This mutation was observed in the homozygous state in the affected offspring of a single Armenian family. The second mutation is a A ⌫ G transition at nucleotide 2080 in which methionine is replaced by valine (M694V). This was observed in a large number of affected individuals bearing four apparently distinct disease associated haplotypes. The third mutation is a T ⌫ C transition at nucleotide 2177 which substitutes alanine for valine (V726A). It was observed in affected individuals bearing the C haplotype in a Druze family and in other FMF patients and carriers bearing this haplotype. An additional mutation in which lysing is replaced by arginine at position 695 (K695R) was observed in an American FMF patient of Northern European ancestry.

Direct sequencing of RT-PCR products or amplified exons from the 8 cDNAs telomeric to v75-1 failed to identify disease-associated mutations.

It is extremely unlikely that the substitutions in v75−1 are actually polymorphisms in tight linkage disequilibrium with "real" mutations on a nearby gene. This hypothesis would require that there be 3 such v75−1 polymorphisms on 3 different haplotypes, each in perfect linkage disequilibrium with the mutations on the "real" FMF gene. While not impossible, such a scenario is at least unnecessarily complex. It is also unclear where such a closely linked gene would be located. The historical recombinants at the 5' (centromeric) end of v75−1 exclude the interval between *D16S3373* and v75−1. On the telomeric side, the 5' end of a novel zinc finger gene is located within 10 kb of the 3' end of v75−1, but thorough screening has revealed no mutations in this later gene (data not shown). Moreover, there are no trapped exons, direct selected cDNAs or expressed sequence tag (EST) hits that map to the interval between them. Finally, and most importantly, the observation of normal chromosomes that bear disease−associated microsatellite and SNP haplotypes but do not have the M680I, M694V or V726A mutations is strong evidence that these are not just haplotype−specific polymorphisms.

Mutation Detection by Fluorescent Sequencing

The entire coding region was sequenced, plus splice cites, in individuals representing seven microsatellite haplotypes. Approximately 100 ng of genomic DNA template was used in PCR reactions to amplify exons and flanking intronic sequences according to the supplier's recommendations for AmpliTaq Gold (Perkin Elmer, Branchburg, NJ) and Advantage−GC Genomic PCR Kit (Clontech, Palo Alto, CA).

The PCR primers were tailed with one of the following sequences:
−21M13 forward: GTA AAA CGA CGG CCA GT; **[SEQ ID NO: 32]**
−28 M13 reverse: CAG GAA ACA GCT ATG ACC AT; **[SEQ ID NO: 33]**
−40 M13 forward: GTT TTC CCA GTC ACG ACG. **[SEQ ID NO: 34]**

After amplification, reactions were run on 1% agarose gels and gel purified using either QIAquick gel extraction kit (QIAGEN, Santa Clarita, CA) or Microcon/Micropure/Gel Nebulizer system (Amicon, Beverly, MA). Alternatively, PCR products were column purified with Microcon−100 (Amicon). Purified amplicons were sequenced with dye primer chemistry (PE Applied Biosystems, or

Amersham, Cleveland, OH). Sequencing reactions were ethanol precipitated and run on an ABI 377 automatated sequencer. Sequence data were analyzed with either Autoassembler 1.4 (PE Applied Biosystems, Branchburg, NJ) or Sequencher 3.0 (Gene Codes Inc., Ann Arbor, MI).

Example 3. Protein Modeling

The deduced amino acid sequence was examined. Two overlapping nuclear targeting signals were detected using the PSORT algorithm (Nakai and Kanehisa, "A knowledge base for predicting protein localization sites in eukaryotic cells," Genomics, 14:897-911 (1992). The first nuclear targeting signal is a four residue pattern composed of a histidine and three lysines. The second is a Robbins/Dingwall consensus (Robbins et al., "Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence," Cell, 615-523 (1991). A bZIP transcription factor basic domain (Shuman et al., "Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP, Science, 249:771-774 (1990) was identified using a PROSITE search (Bairoch et al., "The PROSITE database, its status in 1997," Nucleic Acid Res., 25:217-221 (1997)). The spacing of cystine and histidine residues between residues 375 and 407 (denoted by plus signs in Figure 2) resembles a B-box type zinc finger domain (Reddy et al., "A novel zinc finger coiled-coil domain in a family of nuclear proteins," Trends Biochem. Sci., 17:344-345 (1992)).

Example 4. localizing expression of the protein

The tissues in which v75-1 is expressed are highly consistent with the clinical phenotype for FMF. Based on the nature of the inflammatory infiltrate and the anatomic localization of inflammation in FMF, *MEFV* gene expression might be predicted to be observed in granulocytes and/or serosal cells. Multiple tissue northern blots demonstrated high levels of expression in peripheral blood

leukocytes, primarily in mature granulocytes, but not in lymph nodes, spleen or thymus which are comprised largely of lymphocytes.

Figure 4 shows the expression profile for the v75-1 gene. Figure 4A shows the results of hybridization of a probe derived from exon 2 on multiple tissue Northern blots. A 3.7 kb transcript was found in peripheral blood leukocytes (PBL) and colorectal adenocarcinoma (SW480). The presence of the transcript in peripheral blood leukocytes compare favorably with the symptoms associated with FMF. The detection of the 3.7 transcript in colorectal adenocarcinoma is unexplained.

Figure 4B shows hybridization of the same exon 2 probe on Northern blots with mRNA from purified Polymorphonuclear leukocytes (PMNs) and lymphocytes. PMN lanes represent preparations from different individuals. A β-actin control can be seen at the base of the gel.

The following abbreviations were used in Figure 4: HL-60 (promyelocytic leukemia); K-562 (erythroleukemia); MOLT4 (lymphoblastic leukemia); A549 (lung carcinoma); and G361 (melanoma).

Northern Blot Analysis

To determine transcript size and level of expression in various tissues, multiple tissue Northern blots (Clontech) were hybridized with probes derived from various exons of the gene. These exons were amplified and purified as part of the sequencing protocol for mutation analysis. Larger exons (2, 5, and 10) were labeled by random-priming using Stratagene Prime-It Kit and ^{32}P-dCTP (ICN). Hybridization and washing of blots were essentially as described in Sambrook et al., Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, New York: Cold Spring Harbor Press (1989), except using Hybridisol I (Oncor) prepared hybridization buffer. Hybridization was detected by autoradiography, with 4 hour exposures. Northern blots with mRNA from highly purified peripheral blood lymphocytes, PMNs, and monocytes were the kind gift of Drs. H. Lee Tiffany and Harry Malech.

Example 5. Homologies to other proteins

Figure 6 shows the alignment of the rfp (B30.2) domain of pyrin with homologous proteins. The following abbreviations are used in Figure 6: hum-RFP (RET finger protein; SWISS-PROT P14373); xla-xnf7 (nuclear phosphoprotein xnf7, Xenopus laevis; PIR A43906); pwa-A33 (zinc-binding protein A33, Pleurodeles waltl; SWISS-PROT Q02084); hum-SS-A/Ro (52 kDa RO protein; SWISS-PROT P19474); hum-afp (acid finger protein; GenBank U09825); hum-BT (butyrophilin; GenBank U90552); hum-efp (estrogen-responsive finger protein; PIR A49656); hum-B30-2 (B30-2 gene; PRF 2002339); pig-RFB30 (ring finger protein RFB30, Sus scrofa; EMBL Z97403); hum-Staf-50 (transcription regulator Staf-50; IR A57041).

The invention has been described with reference to various specific and preferred embodiments and techniques. However, it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention. All publications in this specification are indicative of the level of ordinary skill in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated by reference.

What is Claimed is:

1. A nucleic acid sequence encoding pyrin.

2. The nucleic acid sequence of claim 1, comprising the coding sequence of **SEQ ID NO: 2** and variations thereof permitted by genetic code degeneracy.

3. A nucleic acid sequence encoding a familial Mediterranean fever-associated mutant of pyrin.

4. The nucleic acid sequence of claim 3, comprising a mutant pyrin having an amino acid substitution in a rfp (B30.2) domain **[SEQ ID NO: 5]**.

5. The nucleic acid sequence of claim 3, encoding mutant pyrin comprising the amino acid sequence of **SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11 or SEQ ID NO: 13**.

6. A nucleic acid probe or primer comprising at least fifteen consecutive nucleic acids of *MEFV* **[SEQ ID NO: 1]** or a familial Mediterranean fever-associated mutant thereof.

7. The nucleic acid probe of claim 6, wherein the probe hybridizes to *MEFV* under stringent conditions.

8. The nucleic acid probe of claim 6, wherein the probe hybridizes mutant *MEFV* under stringent conditions, the mutant *MEFV* comprising a nucleic acid sequence of **SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12**.

9. The nucleic acid primer of claim 6, wherein the primer amplifies *MEFV*.

10. The nucleic acid primer of claim 6, wherein the primer amplifies a nucleic acid sequence encoding a rfp (B30.2) domain of pyrin.

11. An amino acid sequence comprising **SEQ ID NO: 3**.

12. An amino acid sequence encoding a familial Mediterranean fever-associated mutant of pyrin.

13. The amino acid sequence of claim 12, wherein the mutant comprises one or more amino acid substitutions.

14. The amino acid sequence of claim 12, wherein the mutant comprises an amino acid substitution in a rfp (B30.2) domain.

15. The amino acid sequence of claim 12, wherein the mutant comprises an amino acid substitution in at least one of amino acid residues 680, 694, 695 or 726.

16. The amino acid sequence of claim 12, wherein the mutant comprises an amino acid substitution corresponding to M680I, M694V, K695R, or V726A.

17. The amino acid sequence of claim 12, wherein the mutant comprises **SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11 or SEQ ID NO: 13**.

18. An amino acid sequence encoding pyrin comprising an rfp (B30.2) domain of pyrin **[SEQ ID NO: 5]**.

19. The amino acid sequence of claim 18, comprising an amino acid substitution at residue 680, 694, 695, or 726.

20. The amino acid sequence of claim 19, wherein the substitution comprises M680I, M694V, K695R, or V726A.

21. An anti-pyrin antibody that specifically binds wild type pyrin **[SEQ ID NO: 3]**.

22. The antibody of claim 21, wherein the antibody specifically binds to an epitope in a rfp (B30.2) domain.

23. An anti-pyrin antibody which specifically binds familial Mediterranean fever-associated mutant pyrin.

24. The anti-pyrin antibody of claim 23, wherein the antibody specifically binds to a mutation in a rfp (B30.2) domain.

25. The anti-pyrin antibody of claim 23, wherein the antibody specifically binds to pyrin comprising a mutation at residue 680, 694, 695, or 726.

26. The anti-pyrin antibody of claim 23, wherein the antibody specifically binds to mutant pyrin comprising M680I, M694V, K695R, or V726A.

27. The anti-pyrin antibody of claim 23, wherein the antibody specifically binds to mutant pyrin comprising the amino acid sequence of **SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11 or SEQ ID NO: 13**.

28. A vector comprising a nucleic acid sequence encoding pyrin **[SEQ ID NO: 2]** or a familial Mediterranean fever-associated mutant thereof, operably linked to a functional promoter.

29. A host cell transformed with the vector of claim 30.

30. A kit for diagnostic assay comprising:

a nucleic acid sequence encoding wild-type pyrin; and
at least one nucleic acid sequence encoding a mutant pyrin.

31. A kit for diagnostic assay comprising:
an anti-pyrin antibody which binds wild-type pyrin; and
at least one anti-pyrin antibody which binds mutant pyrin.

32. A kit for diagnostic assay comprising:
at least one pair of primers which amplify a nucleic acid sequence encoding pyrin.

33. The kit of claim 32, wherein the primers amplify a nucleic acid sequence encoding a rfp (B30.2) domain.

34. A method for diagnosing risk of familial Mediterranean fever (FMF), comprising:
analyzing a patient sample for an amino acid sequence of pyrin; and
correlating detection of mutated amino acid sequence with risk of developing FMF.

35. The method of claim 34, wherein analyzing comprises contacting the sample with an anti-pyrin antibody and correlating antibody binding with the presence of pyrin in the sample.

36. A method for diagnosing risk of familial Mediterranean fever (FMF), comprising:
analyzing a patient sample for a nucleic acid sequence encoding pyrin; and
correlating detection of mutated nucleic acid sequence with risk of developing FMF.

37. The method of claim 36, wherein analyzing comprises contacting the patient sample with a labeled nucleic acid sequence encoding wild type or mutant

pyrin and correlating hybridization with the presence of wild type or mutant pyrin.

38. The method of claim 36, wherein analyzing comprises sequencing the nucleic acid sequence of pyrin.

39. The method of claim 36, wherein analyzing comprises sequencing or hybridization of a nucleic acid sequence encoding a rfp (B30.2) domain.

40. A method for producing pyrin in a host cell comprising transforming the host cell with a nucleic acid sequence encoding pyrin.

41. The method of claim 40 wherein the host cell is an animal cell.

42. The method of claim 40 wherein the host cell is a mammalian cell.

43. The method of claim 40 wherein the host cell is a human cell.

44. The method of claim 40 wherein the host cell expresses mutant pyrin prior to transformation.

45. A transgenic animal expressing heterologous wild type pyrin or mutant pyrin.

46. A method for screening compounds for use in FMF therapy comprising: administering candidate compounds to the transgenic animal of claim 45.

47. A method for screening compounds for use in inflammatory disease, comprising administering the compounds to the transgenic animal of claim 45.

FIG. 1

```
1     TATTTTTGTA TTTTAGTAGA GATGGGGTTT ACTGTGTTGG CCAGGCTGGT CTTGTACTCC
61    CAACCTGAGG TGATCCACCC ACCTCGGCCT CCCAAAGTGC TGGGATTACA GGCGTTAGCA
121   CTGTGCCCTG CCCCCAACAT GTAACTTCTG TTAGCTTCAA AGCCACCTCT GGGGCCCTGC
181   ACCACATATG AGCTGAAGGA CACCCGTGCC TTTTCACCCG TGTAGCTCCA GCATCTTGGC
241   ACACTGTCTA GAATGTTCAA TGAATGTGCA CGGAAGAGCA TTCTGGCTCC AGGGAGCGAG
301   GACTGAGTCA GCTCTGGGAA CAGATGAGTC AGGCTGGTGG TCCAGGCATT GCTTTTCAAG
361   TCCTTCATGT GGCTGGAAGA ACCAGTCAAC TGGAACCGGA TCAACAGGGG TGATGGCATG
421   GCAAGAGTTA TCTCCTGGCA GTGCCCTTCT GGCCTCACTT GCCTTCTTGG GCCAGGAAAG
481   GCAAAGCTCA CAGGACTGTA TTCAGTGCCC ACCCCTTCCC CCGTCCTGTG CCATTGGCTC
541   TGGAAGGTCC CTGAAACCCC GAGTCTGGAG GAGAACAGTT GACCAGCAGG GCGGGCCCTC
601   AGCATAGTCC TCTCTGTTCC CACTCACCCG CTCTGCCAGC CCCAGATCCT GGCAGGAAGG
661   AAGATTGGAG GGGGTGTCTG GAATCCAATC CCAGACCTTC CCTTGCAGAC TTGCCCATCT
721   GTCTGTGGTC TAGTGTGGAG GCGAGGTCCA GGGTTTGGGA GGGGTGTGGG GGCACATGTC
781   TGCCAAGGCA TGGAGCCCTC CCAGCTGGAA AATCCTCTGA ACCTGTAAGA AGAGAACACA
841   GCCGGCATGG ACACACCCTT ACCCTTAGTC TCAGTTCCCA CCAAGACACA GAGCATTTCC
901   TGTGCCTTTT CCGCTATTTC ACAACCTGCC TTTTCTTGCT CACCAAGGAC AGAGGCTTCT
961   TTTCCTACCA GAAGCCAGAC AGCTGGCTCG AGCCTCTCCT GCTCAGCACC ATGGCTAAGA
1021  CCCCTAGTGA CCATCTGCTG TCCACCCTGG AGGAGCTGGT GCCCTATGAC TTCGAGAAGT
1081  TCAAGTTCAA GCTGCAGAAC ACCAGTGTGC AGAAGGAGCA CTCCAGGATC CCCCGGAGCC
1141  AGATCCAGAG AGCCAGGCCG GTGAAGATGG CCACTCTGCT GGTCACCTAC TATGGGGAAG
1201  AGTACGCCGT GCAGCTCACC CTGCAGGTCC TGCGGGCCAT CAACCAGCGC CTGCTGGCCG
1261  AGGAGCTCCA CAGGGCAGCC ATTCAGGGTA AGCGGGCCCA GGCCTCCTCC TCATCCAGTG
1321  CTGAGTGCTG GCTGCTTTGT GGGAAAGGGG ACCAGGAGCT CAGAGCAGCT CACTCTGACC
1381  TGGGGATTGG GAGTCTCAGG TCTACCAAAA TCCAGATGAC TTTAGTTCAG GAACGTCCCT
1441  TTCTTCACTC TGGCCTTTGG AACTGGGTTA GTAAACTTCC TTCAGGCTCC TAATGGGTTT
1501  TTTAAGAAGC AGGTCAGGGT CACGAAAGGC AGGAGCTGGA ACACCTGTTC TTTGAGACTT
1561  CTTCACTACA TTTATGATTA ATACTCATGT CAGACAAACA TCTCTAGGTT AGCAAAAAGG
1621  GATTGCTATG CAATCATATG AACGGGGTTG GTATAGAATC TTCTCAGTGC TGTTCACCAT
1681  GTTGGCCAGG CTGGTCTGGA ACTCCTGACC TCAAGTGATC CTCCCGCCTC AGCCTCCCAA
1741  AGTGCTGGGA TTTCAGACAT AGGCCACCGT GCCCGGCTTA TTTTTATTTT TAAAGCGTAT
1801  AATCTGGGTT TTGCTGACCT GTGTAAGATC TTATTTGAAA CAGTTGTCCT GCTTAAAACG
1861  TTTGAAAAGT ACTATTTGAG AAATATAGGC TAGGCATGGT GGCTCACACT TATAAATAAT
1921  CTCAGCACTT TGGGAGGCTA AGGTGGGTGG ATTGCTAGAG CTCAGGAGTT TGAGACCAGC
1981  TTGGGCAACA TGGTGAAACC CTGTCTCTAC CAAAAATACA AAAAAATGAG CCAGGCGTGG
2041  TAGCACACAC CTGTATTTTC AGCTATTGAA AAACAGAAA ACAGGCTGAG GTGAGAGGAT
2101  TGCTTGAGCC TGGGAGGCAG AGGTTGCAGT GAGCTGAGAT CACATCAGGG CAACAGAGCA
2161  AGATCCTGTC TCAAAAAATA AAATAAGAGA GAGAGAAATA CATAGCAACA TCAAGCATGT
2221  TCTTACTGAA TGGTAATTGA CTGCCATTGT CTAGTCTGGG NAGTCCTGAA CTTTTGTTTT
2281  TGAGATGGAG TCTTGCTCTG TCACTCAGGC TGGAGTGCAG TGGCCCGATC TCAGCTCNCT
2341  GCAACCTCCA CATCCCGGGC TCAAGCGATT CTCATGCCTC AGCCTCCCGA GTAGCTGGGA
```

FIG. 1 (CONT.)

2401 CTACAGGTGC GCACCACCGC GTCTGGCTGA GTTTCTTATT TTTAGTAGGA ACGGGGTTTT
2461 GCCATGTTGG CCAGGCTGGT CTCGAACTCC TGACCTCAAA TGATCCTCCC ACCTTGGCCT
2521 CTGGAGAAGC TGGGATTACA GGCATGCGCA CCACGCTCAG CTTATTTTTG TATTTTTAGT
2581 AGAGACGGGG TTTCACCCTG TTGGTCTTGA ACTCCTGATC TCAGGTGATC CTCCCGCCTC
2641 GGCCTCCCAG AGTGCCGGGA ATACAGGCAT GAGCCACCGC GCCCGGCCCG TTGTTTTCCT
2701 CAATTTCTAA ACTTTAATAT CCAAGGGGAT TCTCTCTCCT CTGCCCTGAA TCTTGGGCCC
2761 TAAACGTGGG ACAGCTTCAT CATTTTGCAT CTGGTTGTCC TTCCAGAATA TTCCACACAA
2821 GAAAACGGCA CAGATGATTC CGCAGCGTCC AGCTCCCTGG GGGAGAACAA GCCCAGGAGC
2881 CTGAAGACTC CAGACCACCC CGAGGGGAAC GAGGGGAACG GCCCTCGGCC GTACGGGGGC
2941 GGAGCTGCCA GCCTGCGGTG CAGCCAGCCC GAGGCCGGGA GGGGGCTGTC GAGGAAGCCC
3001 CTGAGCAAAC GCAGAGAGAA GGCCTCGGAG GGCCTGGACG CGCAGGGCAA GCCTCGGACC
3061 CGGAGCCCGG CCCTGCCGGG CGGGAGAAGC CCCGGCCCCT GCAGGCGCT AGAGGGGGGC
3121 CAGGCCGAGG TCCGGCTGCG CAGAAACGCC AGCTCCGCGG GGAGGCTGCA GGGGCTGGCG
3181 GGGGGCGCCC CGGGGCAGAA GGAGTGCAGG CCCTTCGAAG TGTACCTGCC CTCGGGAAAG
3241 ATGCGACCTA GAAGCCTTGA GGTCACCATT TCTACAGGGG AGAAGGCGCC CGCAAATCCA
3301 GAAATTCTCC TGACTCTAGA GGAAAAGACA GCTGCGAATC TGGACTCGGC AACAGAACCC
3361 CGGGCAAGGC CCACTCCGGA TGGAGGGGCA TCTGCGGACC TGAAGGAAGG CCCTGGAAAT
3421 CCAGAACATT CGGTCACCGG TAAATTGTGT TCTTTCCAAC TTTATATCGG CTGCAGAGAA
3481 AGAATGGCTG GCCGGGCACG ATAGCTCATG CCTGTAATCC CAGCGCTTTG GGAGGCCAGG
3541 GCGGGAGGAT TGCTGGAGGC CAAGACTTTG AGACCAGCCT GGTGAATGTA GTGAGACCCC
3601 CGCCATCTCT ATAAACGAAA TTAAAAAAAT AAAAACCCAA AGGTTGGGCA GGGCGTGGTA
3661 GCTCTCGCCT GTAATCCCAG AGCTTTGAGA GGCCTGCACG GGAGGATCTC TTGACCCCAG
3721 GAGTTCCATA CTAGCCTAGG CAACACAGTG AGACCCCATC TCTACAAAAT ACAATAGTGG
3781 CACGCGCCTG TAGTCCCAGC TGCTCGGGTT CACTTGAGCA GACGGAGTTC CAGGCTACAG
3841 TGAGCTGAGG ATCATGCCAC TGCACACCAG CCTGAGCAAC GTAGCCAGAC TCACTTCTAC
3901 AAAACTAAAA AAAAAATTAG CTGGGTATGG TGGCACACGC CTGTAATTCT AGCCACTCAG
3961 GAAGCTGAGG CAGGAGGATT GCTTGAGCCA GGGAGTTCCA GGCTGCAGTG AGCTGAGGAT
4021 GTGCCACTGC ACTCCGGCCT GGGCAACAGA GCAAGACCCT GTCTCTTAAA CATTTTGGGG
4081 GGAAAAAAAA AGAAAGAAAG AATGTCCGAT TGAAAAAGGC AATCAGGTGT TATCAGTGGC
4141 CAAAGAATGG AGAAGGGGAG CTCACCTCTG CAGGCGTCTG CTTGCCAGGG ATGGAGGCA
4201 GGGCGATTTT AGAGTCCAGG GAGGGGAAGG GAGATAGGTA AGCAGGCCCA GGGCAGGGTT
4261 CCATATGTGC AGGCGCTGTC CCCAGCATGC TTCTTCCTAC ATCGCATTCA AACAAACCCT
4321 TCTCCATCTT CTTTAGGGGA GGACCCTTTA GCTTATAACC ATGTGTAAAT GATCCTAAGG
4381 TAACTGGAAG TCACCTCTTC CAGTTTGCAC TGGTTTTGCT CTGATCTTAA CTTCCTCTGG
4441 TTTTTGGCAA GGGATCAGGA GGCTCCAGGC CATCTGGATT TTTTTAAGCA GCTGTCCCTA
4501 TAGGTAAAGA GACTAAAAAA AAACTGTAAA AGAAAAATGC CACCAGTTTA GAGGGTACCG
4561 AGGCTATCCA GGTGACAATT CCATGCTCGT GGTGGGGGCA GCATTCAGAA ACACACTTTC
4621 CTTTTTTTTC CTCCTTTTTT TTTTTGAGAC AGAGTCTCAG TCTGTCTCCC ATGCTGGAGT
4681 GCAGTAGTGT GAGCACAGTT TACTGCAGCC TCAACCTCCT AGGCTCAAGC GATCCTCCCA
4741 CCTCAGCCTT CCAAGTAGCT GAGACTATAG GTGCTCACCA CCACACCTGG TTAATTTTTT

FIG. 1 (CONT.)

4801 TTTTTTTTTT TGTATTTTTT GTAGTTACGA GGACTGTCTA TGTTGCCCAG GCTGGTTTTG
4861 AACTCTTGGG CTCAAGCGAT CCCCCGCCTT AGCCTCTAAA AGTGCTAGGA TTTCAGGTGT
4921 GAGTCACTAC ACCCAGCCTA TGGAACACAC TTTCCAATGC ATTGTTGGCT GGAGAGGAGA
4981 AATCACAGCA CTCAAGGAGG AGAAATAGAA TTGGGGGTCC AGGCCGGGTG CGGTGGCTCA
5041 TACCTGTAAT CCCAGCACTT TGGGAGGCCA ATGGGGCGCG ATCACCTGAG GTGAGGAGTT
5101 CGAGACCAGC CTGCCAACAT GGTGAAACGC CATCTCTACT AAAAATACTA AATTTGCTGG
5161 GCGTGGTGGC GGGTGTCCAT AATCCCAGCT ACTCAGAAGG CTTCGAGGCA GGAGAATTGC
5221 TTGAACCGAG GAGGCAGAGG TTGCAGTGAG CCAAGATCAT GCCACTGCAC TCTAGCCTGG
5281 GCGACAAGAG CAAAACTCTG TCTCAAAAAA AAAAAAAAAA AAGAATTGGG AGTCCAGGGA
5341 CCCCTGAGAC CTGGGAGGGG AAAGGATGTG GTATGCTGCA TGAGTCTTCA AATCCAGAAG
5401 TCCCTGGGTC TTCCAGTGAG AAAGGACCCT GGGATCTGGA AAACCTAGCA TCCTTAGGAA
5461 TAGTGACCTG AAAAGTACTG AAGTATTTCC CCCCTAATTT TCTTTTATCC CTACTGTATT
5521 TTTTTTAATT TTTTTTTTTT TTTAGATATG GGGTCTTGCT ATGTTGCCCA GGTTGGTCTC
5581 GAACTCCTGA TCTCAAACAA TCCTCCCATC TTTGCCTCCG AAACTGCTGG GATTACAGGT
5641 GTGCACCACT GCACCAGGTC CCCACTGTAT TTATATCATT GGGATTCCTG GGTGTCTTCT
5701 AGGGCCGCTT CGTTAATCTG ATGCAGGCTT AGACCCTGAA AAATGCATAT ATGCACAGCT
5761 TCACAAATGT CACATCAAAT TTCAGGTAGT TCTTGGACAC TCTGAAGACC ATCTTTAGAA
5821 TCCAAGGGGT TTATGGACAC CAGGTAGAAA ATCTGGGGAA GACTGGTTAA AAATACTCCC
5881 TCTCACAATA ACCTCACAGC AATGCATCAT CATGGGGTTG AGATTCTACC ATTGCCTTTC
5941 TCTCAGCAGA AAGAAAAGCC TATTGGCTAA AGTCCTAACT ATCTACTGCT GAGGTAGTCA
6001 TTAAAATTAT GTTTGGTTGT GAATAATAGA AACACCCAAA TAACAGTAAC CTCAACAGAA
6061 AAGAAGTTTG TGCCTCCTTC ACATAAATGA TACACAGGCG GTCCCAGGCA GATCCGTGGG
6121 CCAGGACCCT GGGGTCCTGC TGTTGCTCTG TCCCACCAAG TTTGTCCTCA AGCTTCTGCT
6181 CTCAGAAGGT GACGTCCTCA TGCCAGGCAG CAAGATGGAG GAACAGAGGG GAACAGTATC
6241 CCTCGGGAAA GCTCTAGAAG TTTCTAGAAG CTGCTTGTGA CACCTCCATT TACATCCCTT
6301 TGGTCATATT ATTGTCAAAT AGCCACACCT AACTGCAAAG GAGGCTGAGA AATGCAGGGC
6361 ATTTGGGGGG CAATGGGAGG CAGGGAAACA GGGAAACGTG GACAATTAAT TCTATCACGA
6421 GAGAAGGAGG GAGAGTAATT TCTGGTGACT ACTAGCAGTC TCATTTACAG ATGTGCTGTG
6481 AATTTCTGGG ACACTGTGAG GTGGGAGGAG GTAGCAGGGG CTAAAGGATT GAGTGTGTTT
6541 CTATTTCTTT TTTTGTTTTT TTTTTTTTTG AGATGGAGTC TCTCTTGGTC ACCCAGACTG
6601 GAGTGCAGTG GCGCAACTTC AGCTCACTGC AAACTCCGCC TCCCGGGTTC AAGCAATTCT
6661 CCTGCCTCAG CCTCCCGAGT AGCTGGGATT ACAGGTGCCC ACCACCACGT CCGGCTAATT
6721 TTTGTATTTT TAGTAGAGAC AGGGTTTCAC CATCTTGGCC AGGCTGGTCT TGAACTCCTG
6781 ACCTCATGAC CCACCCGCCT CGGCCTCCCA AAGTGCTGGG ATTACAGGCG TGAGCCACTG
6841 CGCTCGGCCT TGTGTTTCTA TTTCTTCTTG TATCTCGTGG CATGTCTGCT TATGAAGTTG
6901 CAATTAGAGT CTTGGAGTAG AGCTATTCAT AACTGTTAGG TCTTCATGAT GAGTTCCAGT
6961 CTTTAGCCCT ATAATGCCCC CCTTCTTTGC TTTTTCTTTT AAGATGGCAT CTTACTCTGT
7021 TGCCCAGGCT GGAGTGCAGT GGTGCAGCAT CAACCTCCTA GGTTCAAGCA ATCCTCCTGT
7081 CTCAGCCTCC CAAGTAGCTG GGATTAGAGG TGTGCACCAC CACACCTGGC TAATTTTTTA
7141 ATTTTTTGTA GAGGTGGGCT CTTGCCATGT TGCCCAGGCT GGTCTCAAAC TCCTGAGCTT

FIG. 1 (CONT.)

7201 AAGCAGTCCT CCCACCTTGG CCTCCCAAAG CACTGGGATT ATAGGCATGA GCCACCACCC
7261 AGCCCCTTCT TTGCTTTCAT TTAATGGTTA TTGAACTCAT ATGTGAGCAG TGGTCTATTT
7321 ATTCCTTCAT TCAATACTCA TTTTCCAAAT GCTTGCATTT GCCAGGTACT CTGCTAGGGG
7381 CTGGGATCCA GCTAGGAGCG AGGTACACAA GTCACCATCC CCTGGAAGCC TCCACTCACG
7441 TTATGGGCAG CCAGGGATGG GTTCAAGTGG CAAAGGAACA CTGGTCAGAA TGTCTCTTTC
7501 CTTGGCATCA CCTGCTAGAT CTATGTCTGT GCAGGAGGAA CAGCACAAGG CCATGGGTCT
7561 TTCTTTAGGA TAAATGCCCA AGAATTCCAA GGCTCAGGAA TGTCTGAGGT CTGGCCCTTA
7621 GCTCTCAGGC CCAGTGGCCT GTTTGCTTCC TCACTGGATG GAAGTCGGGG GAGGACAAGC
7681 TAGGAAGTGG GCAGAGTCTA ACTGAGAACT CGCACATCTC AGGCAAGGGC TGTGTCCGCT
7741 GTGCTTTGTG ATACCTCTGT GTAAGCAACT TGGGTTTGCC ATTCAGGGGG TTTTTCCACT
7801 GCATGTCCCC AGGAAGGCCA CCAGACACGG CTGCGAGTCC CCGCTGCCAC GCCCAGGAAG
7861 GAGACCCAGT TGACGGTACC TGTGTGCGTG ATTCCTGCAG CTTCCCCGAG GCAGTTTCTG
7921 GGCACCCCCA GGCCTCAGGC AGCCGCTCAC CTGGCTGCCC CCGGTGCCAG GACTCCCATG
7981 AAAGGAAGAG CCCGGGAAGC CTAAGCCCCC AGCCCCTGCC ACAGTGTAAG CGCCACCTGA
8041 AGCAGGTCCA GCTGCTCTTC TGTGAGGATC ACGATGAGCC CATCTGCCTC ATCTGCAGTC
8101 TGAGTCAGGA GCACCAAGGC CACCGGGTGC GCCCCATTGA GGAGGTCGCC CTGGAACACA
8161 AGGTAGGCAC TCCCTGCCTG TGGGCTCTTC TCTGCCAGGC ACTTGGACAC ACTGGGCCTT
8221 ACTTCATTTT CCCAACAACT CTGGGTTGTT GGTGCATTAA CCAGCATTCT TGGGCTGGAA
8281 ATGGCAAGAA CACAATATAA ACCAGTCCAG CAAAGAGGGG AGCTACAGGT TTATGTTGCT
8341 CAGAGATCCA GGGGGAGCTG GCTTCAGGTA TGGCTGAATC CAGAGGCTCA GAGGAAGTGC
8401 CTCTCAGCTC TGCTGCCTTT GGCAATTCAG CCATTCCTCC CTCCTCTTTC CTGAGCACCC
8461 CTCCCCATGC CGCTGGCAGC AGCACCCTCA GCCTTGCTAC CAGAAGGAGA TGTTCCCCTC
8521 CAGAGTTGGC ACCAGCTAAA GATGGCAGGA GCCAAATTCA AGCTTTTCAA CAAGTGCTGT
8581 TTTTCCAGAA GAAAATTCAG AAGCAGCTGG AGCATCTGAA GAAGCTGAGA AAATCAGGGG
8641 AGGAGCAGCG ATCCTATGGG GAGGAGAAGG CAGTGAGCTT TCTGGTAAGG TCAGAGGTGG
8701 CTGATGGCCC ATCCGTCCCT GGGAGGAAGG TGGGAAGAGT GAGCAGGGGT CCCCGAGATT
8761 CTGCTGTGGT TCACAGGGCA GCAGGGATGG CCACCTCCTC TCAGGGGACA GAGGGTAACC
8821 AGCAGCCAAG GGTAAGCTCA TCCCTGTAGA GGGAGACCAC CCCCAGCAGG CAGGGGTCAC
8881 CTCTGAGGAT CCTGTCATGC TTTCTCATAC TCACCAGAAG ATGGTAGAGA GCAACCTATG
8941 CCGGTGACTA CTGCAGAAAG ATGGGATTGA GGAAAAGGGA GGAGAACGCC ACTTTCTTTT
9001 TTTGTGACGG AGTCTCGCTC TGTCACCCAG GTTGTAGTGC AGTGGTGTGA TCTTGGCTCA
9061 CTGCAACCTC TGCCTCCCGG GTTCAAGCGA TTCTCCTGCC TCAGCCTCCT GAGTAGCTGG
9121 GATTATAGGT GAGTGCCACC ATGCCTGGCT AATTTTTGTA GTTTTAGTAG AGATGGGGTT
9181 TCACCATGTT GGTCAGGCTG TTCTCGAACT CCTGAACTCG TGATCCGCCC GCCTTGGCCT
9241 CCCAAAGTAC TGGGATTACA GATGTGAGCC ACTGCGCCCG GCCAAGAACA CTTTTAACTT
9301 CATAATTTAC TCTCTGTTTT TTTGTTTTGT TTCCAAGATG GAGTCTCGCT CTGTCACCCA
9361 GGCTGGAGTA CAGTGGCACG ATCTTGGCTT GCTCCAACCT CCACCTCCGA GGTTCAAGCA
9421 ATTCTCCTGC CTCAGCCTCC TTAGTGGCTG GAATTACAGG CGCCTGCCAC CGCGCCTGGC
9481 TAATTTTTGT ATTTTTAGTA GAGACGGGAT TTCACCGTGT TGGCCAGGCT GGTCTCAAAC
9541 TCCTGACCTC AGGTGATCCA CCTGCCTCGG CCTCCCAAAG TGCTGGGATT ACAGGTGTGA

FIG. 1 (CONT.)

9601 GCCATCGTGC CTGGGCTGGT TTTTTTGTTT TTTAGGGTTT TTTTTTTTTT TTTTTTTTGA
9661 GATGGAATCT CACTCCGTCG TCCAGGCTGG GGTGCAGTGG TGCAATCTCG GCTCACTGCA
9721 AACCTTCGCC TCCCCAGTTG AAGCAATTCT CCTGCCTCAG CCTCCCGAGT TGCTGGGACT
9781 GTAGGCACAT GCCACCACTC CTGGCTAATT TTTGTATTTT TAGTAAAGAC AGAGTTTCCC
9841 CATGTTGGCC AGGCTGGTCT CGAACTCCTG ATCTCAAGTG ATCTGCCCAA CTCAGCCTCC
9901 CAAAGTGCTG GGATTACAGA CATGAGCCAA TGCACCCAGC CCAAATTTCC CCATTTTATA
9961 AGACAACATT TATATTGGAT TAGGGACCCA CCCAATCCCA GTAGGACCAC ATCTTAACTA
10021 ATTACATCTG CAAGAACTCT TATCTCCAAA TAAGATCACA TGCTGAGTAC TGGGGGTTAG
10081 GGCTTCAACG TGTAAATTTT GGAAGGGACA CAGTTAAACC TTAACACCAG GTTTAAGGAC
10141 ATTTTCCCAG AGCTAGCCCC AGCCATGCTC AGTCTTTTCT GGAAGGTTCC AGACAATATC
10201 GCCTCCTGCT CTGGAATCTA GGCCTTGAAG AGGCAGCATA AGCCCACCTC TTATCCACCT
10261 CCAGGAGGTG GGCTTCTGGG GGTTCCTGGA CATCCAGGTC CACCCACAGC ACAGACCCCC
10321 ATACCTCCCT GTCCTCTGCT CCCCAGAAAC AAACTGAAGC GCTGAAGCAG CGGGTGCAGA
10381 GGAAGCTGGA GCAGGTGTAC TACTTCCTGG AGCAGCAAGA GCATTTCTTT GTGGCCTCAC
10441 TGGAGGACGT GGGCCAGATG GTTGGGCAGA TCAGGAAGGC ATATGACACC CGCGTATCCC
10501 AGGACATCGC CCTGCTCGAT GCGCTGATTG GGGAACTGGA GGCCAAGGAG TGCCAGTCAG
10561 AATGGGAACT TCTGCAGGTG GGTGTGCCTG GGCCCGGCTT TCTTGGGTCC CCTGTGCCTA
10621 TCAGGATGCC TCAGGCTCCC AGCTCTGCCA TCAGCCGTGC TGGAACAAGT GGGTGAAGCC
10681 CTAAGGCCTA GGATAGGACT TGGTCTTGGT GACCCACAGT GCCTCTTGTG CCCAGACCCC
10741 TTTGATGAGG TCTCTCAGGA GCCCAGGGTG GCCTGGTATC CAGGGGATCT CTGCCATTTC
10801 CCAGAAGGGA TCAGCAGGGC TTGAGGGCCG TTCCATTGCA GGCCTCGCCA CCTGGGATGC
10861 CTGAATTCCC GTGGTTAGAA TTAGACTTGA AGAAAGGTGC TCCACTTCCA CTGACACCCT
10921 AGGGCAGGGA GCCCTGGTAA GTGCAGCGGG GAGCTAAAAG TCCAGGAGCC CAGAAGTAGA
10981 GGCCAGGAGT CAGCCCAGCC ACTAGGAGCC TGGTAACCGA CAGTTTCCTT CTTTTTTCTC
11041 CTAGGACATT GGAGACATCT TGCACAGGTA CAGCGAGGTC CTGTGGTGTA CCCTGGGGTG
11101 TCTTGCAGAA AGCATATGGG GGAGACAGTC CCAGAAGGGA CCTGGGAGGG AGATGTTCCC
11161 AACCCCGGGG TCTGTGATTC CAGACTCCTC CTTTTTTCTG CAGCTTCCCA AAGCCTCTCT
11221 GGATTTGATA GGGAGAAGGG CATCTGGTCA GCAGGGAGGC TGGCCGGGTA TGGAGCTGCA
11281 GACTGGGAAG GGTGAATTCA GCCCATCCTG CTGAAACAAG ATGGAGGCTC CCTAAGAAAC
11341 CTTCCGAGTG CATTGTGTCC CGTGCAGTTC ATCTGATGAA AGCTGCCCCT TCAGGCCTAC
11401 TGGTGGCCTT GGGAAGCTTG TTTGGAGTGG AGCTGGGCTA AGCCCAGCAG GAAGGGGAGG
11461 GGAGGGAAGG GACAGGAAGA GGCTAAGCCT TAAAATCACC TGGGAGCTTT ACAAAATCCC
11521 GGTGTCCTTT TGTGTCTGGC TTCTTCACTT AGCATAATGT CTTCGGGCTT CATCCGTGTT
11581 GTAACGTGTA TCAGAATTTA TTTTCTTTTT ATGGCTGAAT CATAGTCCAG TGTGTGTTCA
11641 TACATTTTGC TTATCCATTC ATGGATATCG GGACTTCTTC TAACTTTTGG TTTGTGAATA
11701 ATGTTGCTAT GAACAAGGGT GTACAAATAT CTGCTTGAGA CCCTGCTTTG TTATTTTGGG
11761 TACCTACCCA GAAGTGGAAC TGCGGGACCA TGTGGTTATC CTGTGTTTAA TTTTTTTTGA
11821 GGAACCACCA TCCTAATTCT CACAGGGGCT GCATCGCTTC ACATTCCCAC CAGCAGCACA
11881 CAGGGGCTCC AGTTTCTCCA CATCTTTGCC ATCACTTATT TTCTTCTGTT TCACTCTCTC
11941 TCTCTCTCTT TTTTTTTGAA GACAGCGTCT TGCTCTGTCA TCCAGGCTGG AGTGCAGTGG

FIG. 1 (CONT.)

12001 CGCGATCTTG GCTCACTACA ACCTCTGCCT CCCAGGTTCA AGGGATTCTC CCACCTCAGC
12061 CTCCCTAGTA GCTGGGACTA CAGGAGCGTG CCACCATGCC CAGCTAATTT TTTTGGTAGA
12121 CAGGGTTTCA CCATATTAGC CAGGCTGGTC TCAAACTCCT GACCTCAAGT GATCCACCCA
12181 CCTTGGCCTC CCAAAGCGCT GGGATTGCAG GCGTGAGCAC CGTGCCCAGC CATTTCTCTT
12241 TCCTTCCTTC CCTCCCTCCC TCCCTTCCTT CCTTTCTTCC TTCCTTCCTT TCTTTTCTTC
12301 TTGAGACAAG GTCTCACTCC CATCACTAAG GCTGGAGAGC AGTGGCACAG TCACAGCTCA
12361 CTGCAGGCTC AGCTTCCTGG GCTCGGGTGA TTCTGAGTAG CTGGCATCCT GAGTAGCTGG
12421 GACTACAGGC ATGTGCTACC ACTTCCGGCT ACTTTTTTGT ATTTTTAATA GAGACAGGGT
12481 TTCGCCATGT TGCCCAAGCT GGACTTGAAC TCCTGGGCTC AAGCGATCCC ACTGCCCCGG
12541 CCTCCTGAAG TGCTAGGATT ACAGGCATGA GCCACCATAC CTGGTCTATT TTTTTCTGTT
12601 GTTGCTGTTT TTATAATAGC CATTCTAATG GATGTGAAGG GATATTTTGT TGTGTGTGTT
12661 TTTTTTTCAT TTATTATCTT TTTATTTCAA TAGAAAGAAA GGGGTGTATA ATCAATTTGA
12721 CATAGATAAT TCTAGTAGAT AATATCAATG TCATTTTAAG TCCATTCTGA AAACTCCTTG
12781 TGGTTTTGAT ATCCATGTCT TTAAAGCACC CCAGTACATG ACAGTCTGTG GCCAAAGTTG
12841 AGGACCAGCA TTTAGACCTC TGAATCCAGG GAAGACTTTT CTTTGTGTAG CTCAGGCTGG
12901 GCTAGGTGTG CCTTGTGGAG AATGTAGTTC ATTTCCAGCT CACGGGTACT TGGGCCACCC
12961 CCTCGCTCCG GCCTTCTCTG GTCAACAGTC TTTTGTCTCT AGGGCTAAGA CATGGCCTGT
13021 SCCTGCAAAG TGGACCACTC CTCAAGAGAT AAAACAAAAG ATCCAACTCC TCCACCAGAA
13081 GTCAGAGTTT GTGGAGAAGA GCACAAAGTA CTTCTCAGGT AGATGGGCTT GGGAGAAGAT
13141 TGGAGGTGCA TGCTCACTTC CTCCCTAAGA TCCACATAGC CCAGAGCCCC TCACTTCCCT
13201 CCTCTTCCCC TGGTCTTGCT GACCTGCCTT CAACCTCTCC TCCATCTGTC CCTGGCTGAG
13261 GGACCTAACT CCAGCTTCTC TCTGCTCCCT TTCCCACATT TTAGAAACCC TGCGTTCAGA
13321 AATGGAAATG TTCAATGGTG AGTCCAGCGG TAATGGTGTG TGCTGGCCTG GGGTTGTTGC
13381 AGTGTTCCCT TGTGCTGTTG ACTTGAGGGG CCCTATTTAG AAGACAAAAA AAAAAACCAA
13441 ACACCTGGAG CAAAGGTAGG AGAAAGGTCA TGGCAGGCCC CCCAGGCTCT GTGCGTGACT
13501 CATTGACTGA GTTGACTCAT TAGACCACAG TCCCCAACAT GGCCTGGGTT CCTGGGAGGA
13561 ACGGGATTAT ACCCAACATA GCATGCAGGG CCCTAAGCAG GGGGTTCCTT GTCTTTCCTT
13621 GTTGTCAGGA CAGTGTAATT TAGCCCCTCT TAATGCTAAT GCTCAGGAAT TTTTTCCCTA
13681 TCTGATTTTT CTCCGTAGTT CCAGAGCTGA TTGGCGCTCA GGCACATGCT GGTAAGTGCC
13741 CAGATCAAGG CAAGTGGCCC TGGCCTGCTG GATCCCTGTG CTCTCCCCTA CCACGTTCCA
13801 GAAGAACTAC CCTGTCCCTG TTTCCTGCAG GTGGGGAGAA CCCTGTAGGG ATGTTGCCCA
13861 TGGACCCCTA CCTAGGTATT CAAATTTTCT TTGCAGTTAA TGTGATTCTG GATGCAGAAA
13921 CCGCTTACCC CAACCTCATC TTCTCTGATG ATCTGAAGAG TGTTAGACTT GGAAACAAGT
13981 GGGAGAGGCT GCCTGATGGC CCGCAAAGAT TTGACAGCTG TATCATTGTT CTGGGCTCTC
14041 CGAGTTTCCT CTCTGGCCGC CGTTACTGGG AGGTGGAGGT TGGAGACAAG ACAGCATGGA
14101 TCCTGGGAGC CTGCAAGACA TCCATAAGCA GGAAAGGGAA CATGACTCTG TCGCCAGAGA
14161 ATGGCTACTG GGTGGTGATA ATGATGAAGG AAAATGAGTA CCAGGCGTCC AGCGTTCCCC
14221 CGACCCGCCT GCTAATAAAG GAGCCTCCCA AGCGTGTGGG CATCTTCGTG GACTACAGAG
14281 TTGGAAGCAT CTCCTTTTAC AATGTGACAG CCRGATCCCA CATCTATACA TTCGCCAGCT
14341 GCTCTTTCTC TGGGCCCCTT CAACCTATCT TCAGCCCTGG GACACGTGAT GGAGGGAAGA

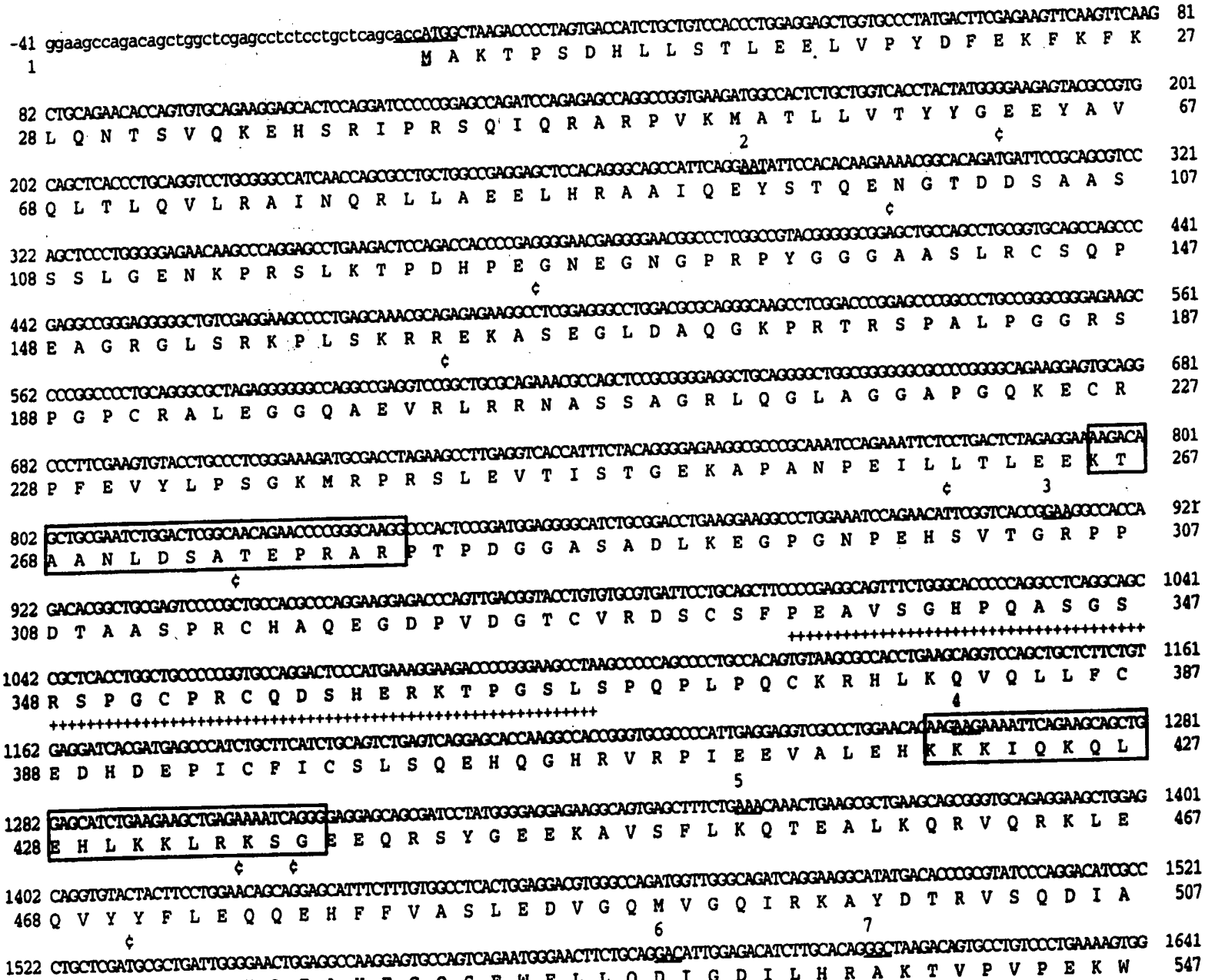
FIG. 1 (CONT.)

14401 ACACAGCTCC TCTGACTATC TGTCCAGTGG GTGGTCAGGG GCCTGACTGA ATGCCCAACA
14461 CTGCATCTCT CTTCCTGCTT CTGGCCTTGT ATCTTGCATT CACACTCAAT AGTCACGGAA
14521 TGCCGACTAG GTGCTAGCTG CTATGGGAAA TGCMAAAATA ACAAAATAGT TACTGTGCCC
14581 ACGGAGCCCT ACCCGATTAT AGCAGAGGTA AGTTAGGAAC GAACATGTTA GTCAATCCGG
14641 GTGAAGACAT GTACTGATGA CACACCATGG ATTTCAGAGG AGGAAGTACG GAGTCGTTGC
14701 ATAATCCGCC CCTGGTGGGT GGCACTCTCA GGTGCTCCTG AACAGAAGAT TTGGCCCTCA
14761 TTTTCCCTCA GAACCCCACG GCAAGGATAT ATGTCCCCTT GTTCTCTCTG CTTCTGTCTT
14821 GAGGATATGG GAAGCCTAGA GAAACGCAAG CAGACTGGAT TGGGATAGAA GTATTTGTGT
14881 ACCTGGATTA ATGAACTATG ATTTTTTTTT TTTTTTTTTG AGACCAAATC TTGCTCTGTG
14941 GCCCAGGCTG GAGTGCAGTG GCACGATCTC AGCTCACTGC AACCTCCACC TCCCAGGTTC
15001 AAGCGATTCT CCTGCCTCAG CCTCCTGAGC AGCTGGGGAT TACAGGTGCG TGCCACCACA
15061 CCAGGCTGGT TTTCTTGTAT TTTTAGTAGA GACGGGGGTT TCACCATGTT AGCCAGGCTG
15121 GTCTCGAACT CCTGACCTCA GGTGATCCAC CCGCCTCAGC CTCCCAAAGT GCTGGGATTA
15181 CAGGCATGAG CCACTGTGCC CGGCCTATGA TTCTTTTTTT TTTTTTTTTT TGAGACAAAG
15241 TTTTGCTCTT GTCACCCAGG CTGGAGTGCA GTGGTGCAAT CTTGGCTCGC AACCTCCGCC
15301 TCCCAGGTTC AAGAGATTCT CCTGCCTCAG CCTCCGAAGT AGCTGGGATT ACAGGCGCCC
15361 GCCACCATGC CCGGCTAATT TTTTGCATTT TTAGTAGACA TGAGGTTTCA TCATGTTGGC
15421 CAGGCCGGTC TCAAACTCCT GACCTCAGGT GATGCACCCA CCTCAGCCTC CCAAAGTGCA
15481 GGGATTACAG GCATGAGCCA CCATGCCGGG CCATGATTCT TAAGAGAATT GACTGGGCCT
15541 CATGAATAAA AAAATTAGAA AATCTGGTCA TTTGCATTTG TCACTCAATC ACTGTGGAAT
15601 CCCATTTCCC GACTGCATTT NCAGGAAGTC AGATGGGACT ACTGTCATGG AAAAACATTT
15661 GGGCATGTTA TTTCCAAGTG TCAGATTATT CTGTCTTGGT TTGTATGGGA AAATCTGCGG
15721 GTTGTGGAAT ATTAGGTTCT ACTTCACACA CATCCCGTGC ATTTGTCCTT CATTTAAAGA
15781 GATGTAAAGG GGCCGGGCAT GGTGACTCAC ATCTGTAATC TCAGCATTTT GGGAGGCAAA
15841 GGCGGGTGGA TCGCCTGAGC CCAGGGATTG AGACCAGCTG GGCAATGTGG CGAAAACCCG
15901 TCTCTACAAA AAATACAAAA ATTAGCCATA GGGATGGGGG TGGGAGGATG GCTTGAGCGC
15961 AGGAGATCGA GGCTGCAGCA GTGAACTGAG ACTGCACTAC GGCAATCCAG CCTGGGCAAC
16021 AGAGTGAGTC CCTGTCTCCA AAAAGTGGAT GTTAGGAGTA CAAAAATCAA ATGAAGATTA
16081 GATCCAAACT CCTATGCCAA CTCCTCTGTC TTCACTACTA GAGTGTAGAT TAGACTCAGA
16141 TACTCCATGG CTATGATGAG AGCAGGTAAA CTTGCTGGGC TTTCCTCCAC GAGTTTTATT
16201 CTATAAGAGT AATCCACATC CCAGGACAGT TCACATGACC TACGGCTTAG CTGTTCCCTG
16261 CGGTGGGTCA TGTCTTATTC CCGATTCTCC CTTGTTATAA GCTTTTCATG AATATCTTTG
16321 TGTATATTTT CCACCACCTC ACCATATACA TATTTTTTTC TCCTGTGTTA TTCCTAAAAT
16381 GGTTCCTGAA TGTGAAATAT CTGATAATGC TTCCTACGGG TTGCCATACC ATCCTTTGCA
16441 AAGATTTTTA AAATATTTCA TGCCCAAAGC AATGACTGCC ATTTAAAATT TTTTGCTGA
16501 TTTAATAGGG ATGTAATGAG GCCTTACTTC TGTTTTATTT CATTACCTGT TAATGAGGCT
16561 GTGAATTTTT CCATGTGAAT TTCTGCTTTT TGCTTCATTC TATGGAAATT GTACAGTTCC
16621 TTTGAATACT TGCTATTTGG AATCTACATA TTGAATTTCG TGTTTTGCTG TACTTCCTCA
16681 TTACATGGTT TTAGGCTGGG TGCGGTGCTC ACGCCTGAAA TCCCAACATT TTGGGAGCCG
16741 GAGGTGGGCA GGATCGGTTG GCAATCGAGG GTTTCGAGAC CGAGCCTGGG CAGACATGGC

FIG. 1 (CONT.)

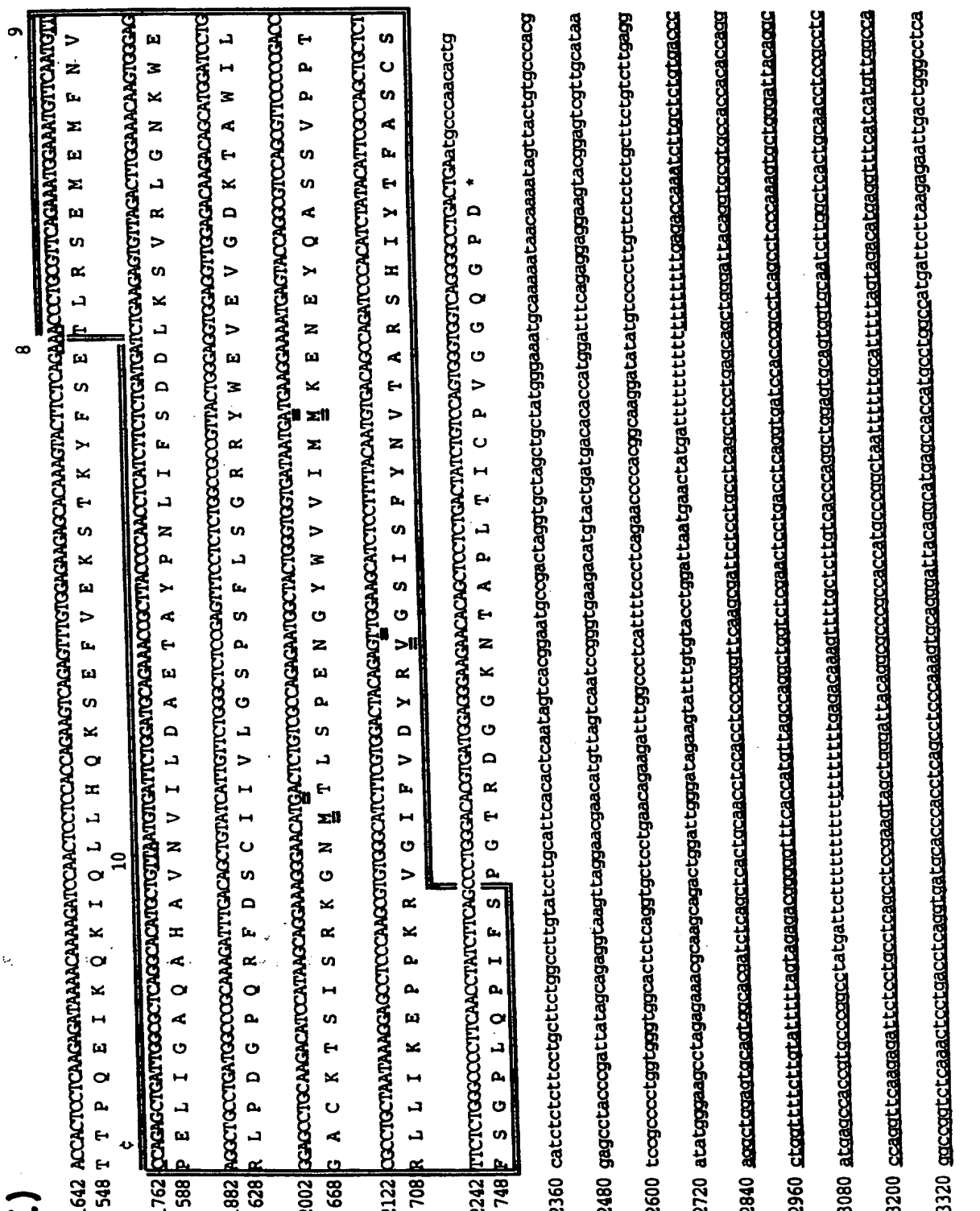
16801 GAAACCTCGC CCTCTACCTA GAAAGATAAA CAAATTAGCG CAGGCAATGG TGGTGAGCAC
16861 CTGTAGTCCT AGCTGATAAG GTCTAGGTTG A

FIG. 2



SUBSTITUTE SHEET (RULE 26)

FIG. 2 (CONT.)



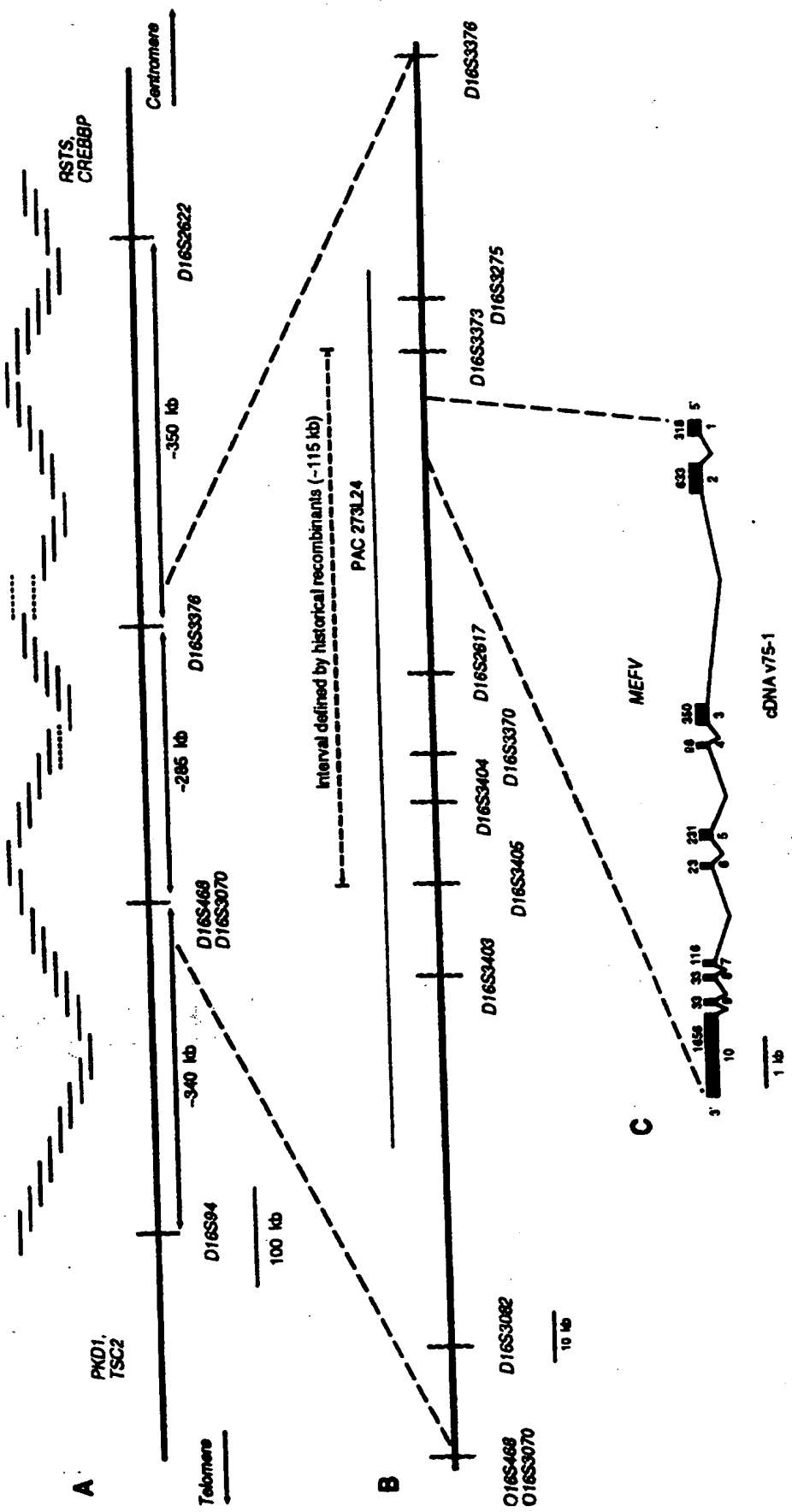

A
PKD1, TSC2
Telomere
D16S94
D16S468
D16S3070
-340 kb
-285 kb
D16S3376
-350 kb
D16S2622
RSTS, CREBBP
Centromere
100 kb
B
D16S3082
D16S3403
D16S3405
D16S3404
D16S3370
D16S2617
D16S3376
D16S3373
D16S3275
PAC 273L24
Interval defined by historical recombinants (~115 kb)
10 kb
C
MEFV
cDNA v75-1
1 kb

FIG.3

FIG. 4 A

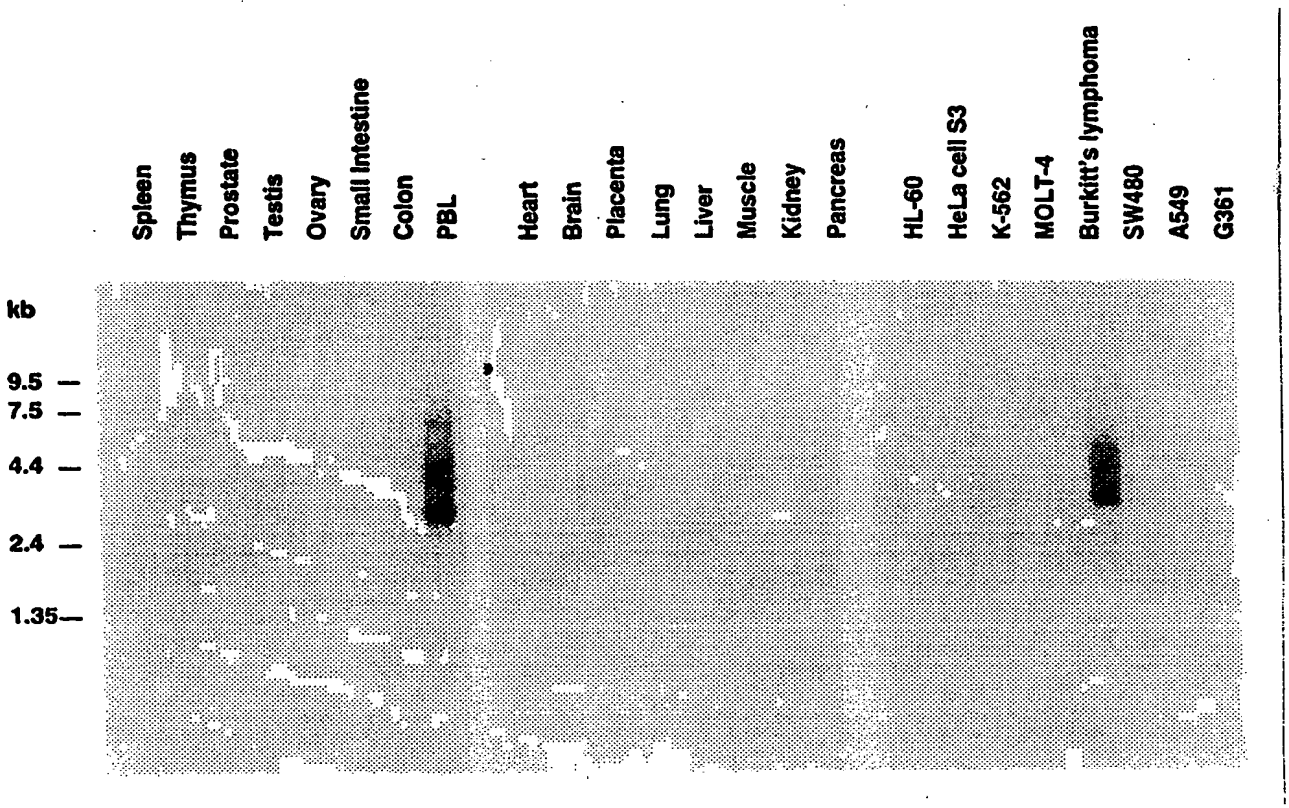
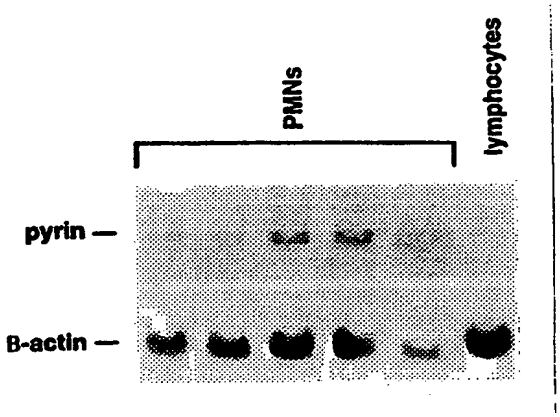


FIG. 4 B



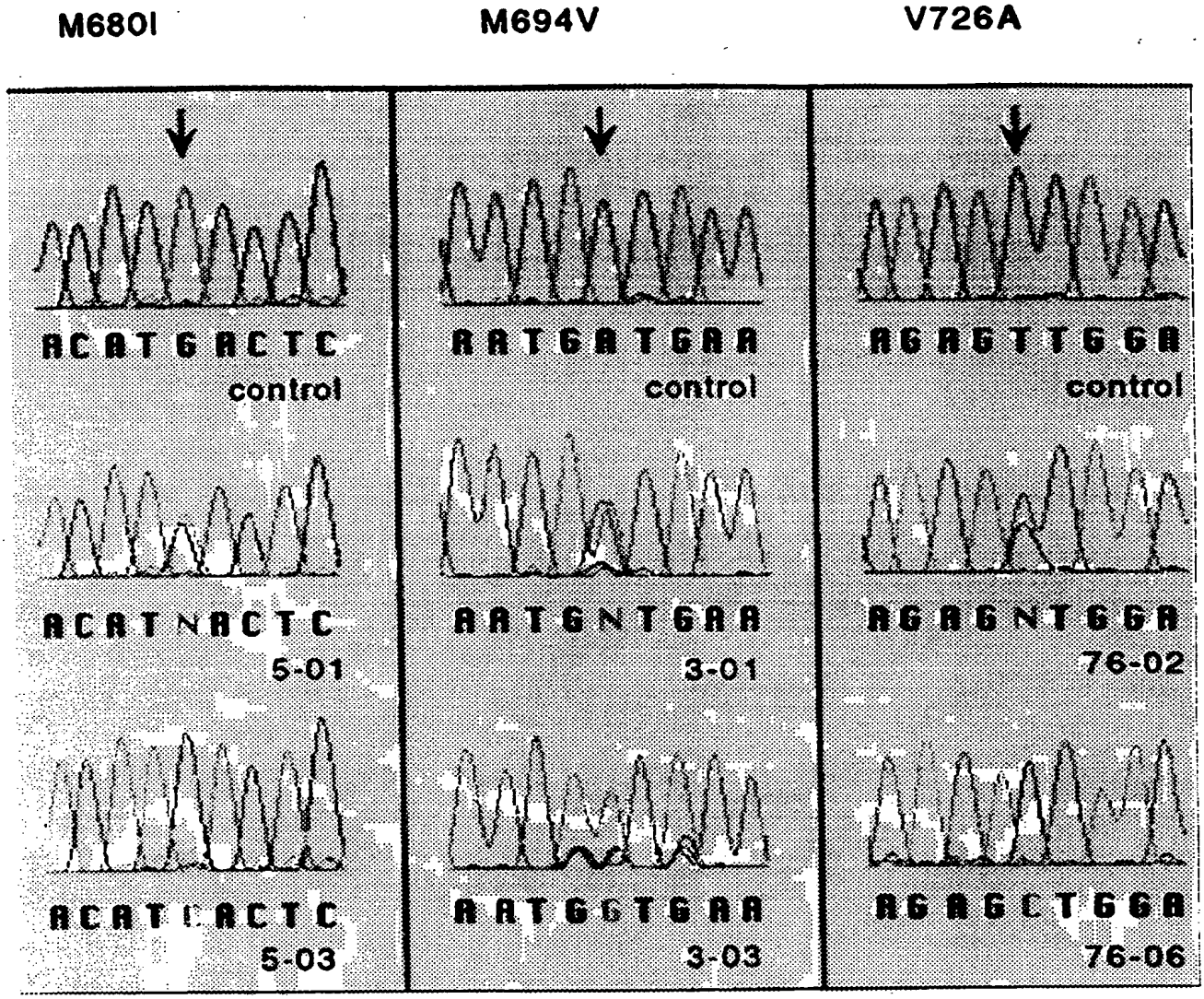


FIG. 5

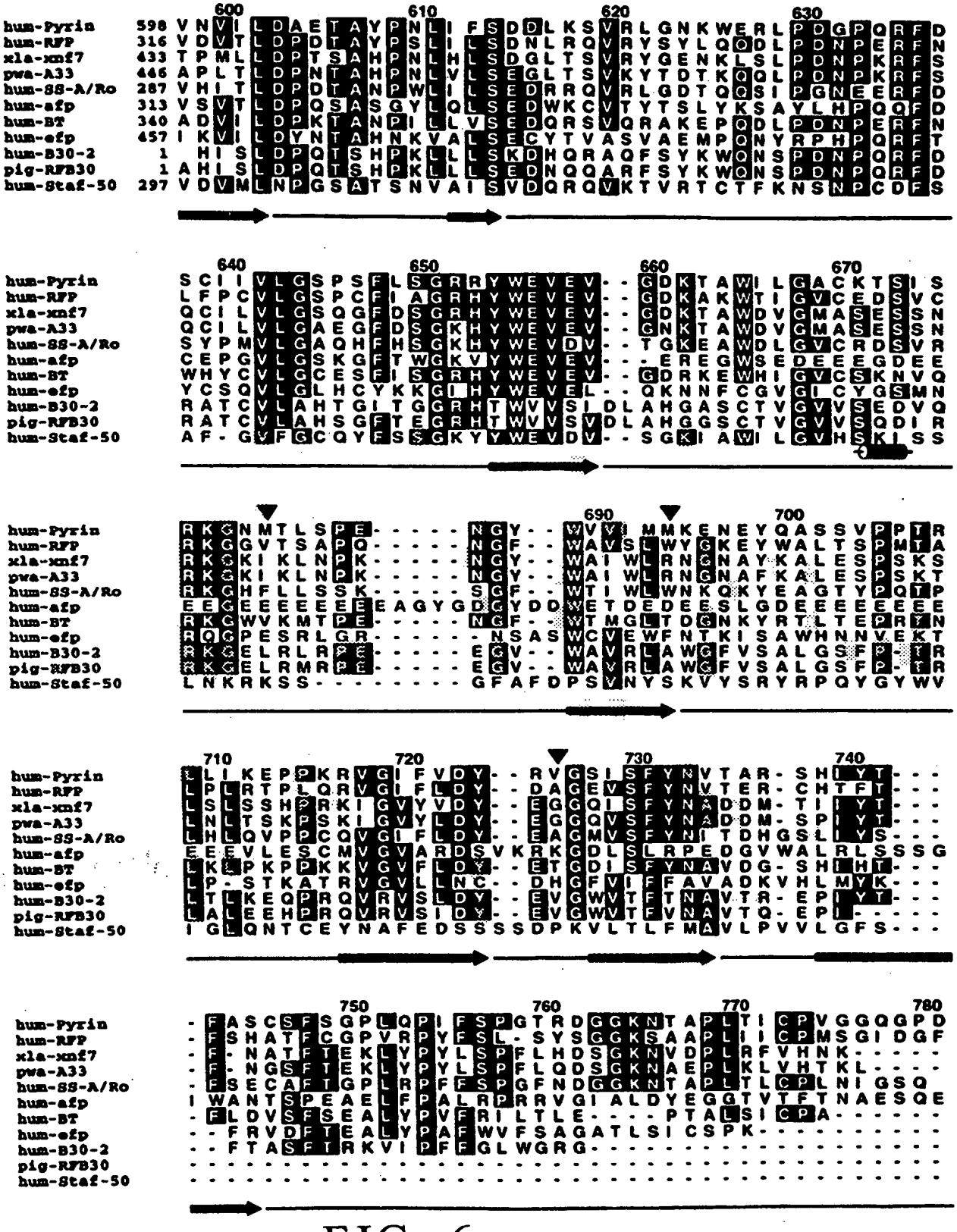

600 610 620 630
hum-Pyrin 598 V N V I L D A E T A Y P N L I F S D D L K S V R L G N K W E R L P D G P Q R F D
hum-RFP 316 V D V T L D P D T A Y P S L I L S D N L R Q V R Y S Y L Q Q D L P D N P E R F N
xla-xnf7 433 T P M L L D P T S A H P N L H L S D G L T S V R Y G E N K L S L P D N P K R F N
pwa-A33 446 A P L T L D P N T A H P N L V L S E G L T S V K Y T D T K Q Q L P D N P K R F S
hum-SS-A/Ro 287 V H I T L D P D T A N P W L I L S E D R R Q V R L G D T Q Q S I P G N E E R F D
hum-afp 313 V S V T L D P Q S A S G Y L Q L S E D W K C V T Y T S L Y K S A Y L H P Q Q F D
hum-BT 340 A D V I L D P K T A N P I L L V S E D Q R S V Q R A K E P Q D L P D N P E R F N
hum-efp 457 I K V I L D Y N T A H N K V A L S E C Y T V A S V A E M P Q N Y R P H P Q R F T
hum-B30-2 1 H I S L D P Q T S H P K L L L S K D H Q Q R A Q F S Y K W Q N S P D N P Q R F D
pig-RFB30 1 A H I S L D P Q T S H P K L L L S E D N Q Q A R F S Y K W Q N S P D N P Q R F D
hum-Staf-50 297 V D V M L N P G S A T S N V A I S V D Q R Q V K T V R T C T F K N S N P C D F S
640 650 660 670
hum-Pyrin S C I I V L G S P S F L S G R R Y W E V E V - - G D K T A W I L G A C K T S I S
hum-RFP L F P C V L G S P C F I A G R H Y W E V E V - - G D K A K W T I G V C E D S V C
xla-xnf7 Q C I L V L G S Q G F D S G R H Y W E V E V - - G D K T A W D V G M A S E S S N
pwa-A33 Q C I L V L G A E G F D S G K H Y W E V E V - - G N K T A W D V G M A S E S S N
hum-SS-A/Ro S Y P M V L G A Q H F H S G K H Y W E V D V - - T G K E A W D L G V C R D S V R
hum-afp C E P G V L G S K G F T W G K V Y W E V E V - - E R E G W S E D E E E G D E E
hum-BT W H Y C V L G C E S F I S G R H Y W E V E V - - G D R K E W H I G V C S K N V Q
hum-efp Y C S Q V L G L H C Y K K G I H Y W E V E L - - Q K N N F C G V G I C Y G S M N
hum-B30-2 R A T C V L A H T G I T G G R H T W V V S I D L A H G A S C T V G V V S E D V Q
pig-RFB30 R A T C V L A H S G F T E G R H T W V V S V D L A H G G S C T V G V V S Q D I R
hum-Staf-50 A F - G V F G C Q Y F S S G K Y Y W E V D V - - S G K I A W I L G V H S K I S S
690 700
hum-Pyrin R K G N M T L S P E - - - - - N G Y - - W V Y I M M K E N E Y Q A S S V P P T R
hum-RFP R K G G V T S A P Q - - - - - N G F - - W A V S L W Y G K E Y W A L T S P M T A
xla-xnf7 R K G K I K L N P K - - - - - N G Y - - W A I W L R N G N A Y K A L E S P S K S T
pwa-A33 R K G K I K L N P K - - - - - N G Y - - W A I W L R N G N A F K A L E S P S K T
hum-SS-A/Ro R K G H F L L S S K - - - - - S G F - - W T I W L W N K Q K Y E A G T Y P Q T P
hum-afp E E G E E E E E E E A G Y G D G Y D D W E T D E D E E S L G D E E E E E E E E
hum-BT R K G W V K M T P E - - - - - N G F - - W T M G L T D G N K Y R T L T E P R T N
hum-efp R Q G P E S R L G R - - - - - N S A S W C V E W F N T K I S A W H N N V E K T
hum-B30-2 R K G E L R L R P E - - - - - E G V - - W A V R L A W G F V S A L G S F P - T R
pig-RFB30 R K G E L R M R P E - - - - - E G V - - W A V R L A W G F V S A L G S F P - T R
hum-Staf-50 L N K R K S S - - - - - - - - G F A F D P S V N Y S K V Y S R Y R P Q Y G Y W V
710 720 730 740
hum-Pyrin L L I K E P P K R V G I F V D Y - - R V G S I S F Y N V T A R - S H I Y T - - -
hum-RFP L P L R T P L Q R V G I F L D Y - - D A G E V S F Y N V T E R - C H T F T - - -
xla-xnf7 L S L S S H P R K I G V Y V D Y - - E G G Q I S F Y N A D D M - T I I Y T - - -
pwa-A33 L N L T S K P S K I G V Y L D Y - - E G G Q V S F Y N A D D M - S P I Y T - - -
hum-SS-A/Ro L H L Q V P P C Q V G I F L D Y - - E A G M V S F Y N I T D H G S L I Y S - - -
hum-afp E E E V L E S C M V G V A R D S V K R K G D L S L R P E D G V W A L R L S S S S G
hum-BT L K L P K P P K K V G V F L D Y - - E T G D I S F Y N A V D G - S H I H T - - -
hum-efp L P - S T K A T R V G V L L N C - - D H G F V I F F A V A D K V H L M Y K - - -
hum-B30-2 L T L K E Q P R Q V R V S I D Y - - E V G W V T F T N A V T Q - E P I Y T - - -
pig-RFB30 L A L E E H P R Q V R V S I D Y - - E V G W V T F V N A V T Q - E P I - - - - -
hum-Staf-50 I G L Q N T C E Y N A F E D S S S S D P K V L T L F M A V L P V V L G F S - - -
750 760 770 780
hum-Pyrin - F A S C S F S G P L Q P I F S P G T R D G G K N T A P L T I C P V G G Q G P D
hum-RFP - F S H A T F C G P V R P Y F S L - S Y S G G K S A A P L I I C P M S G I D G F
xla-xnf7 - F - N A T F T E K L Y P Y L S P F L H D S G K N V D P L R F V H N K - - - - -
pwa-A33 - F - N G S F T E K L Y P Y L S P F L Q D S G K N A E P L K L V H T K L - - - -
hum-SS-A/Ro - F S E C A F T G P L R P F F S P G F N D G G K N T A P L T L C P L N I G S Q
hum-afp I W A N T S P E A E L F P A L R P R R V G I A L D Y E G G T V T F T N A E S Q E
hum-BT - F L D V S F S E A L Y P V F R I L T L E - - - - P T A L S I C P A - - - - - -
hum-efp - - F R V D F T E A L Y P A F W V F S A G A T L S I C S P K - - - - - - - - - -
hum-B30-2 - - F T A S F T R K V I P F F G L W G R G - - - - - - - - - - - - - - - - - -
pig-RFB30 -
hum-Staf-50 -

FIG. 6

INTERNATIONAL SEARCH REPORT

Int. onal Application No
PCT/US 98/17255

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47 C07K16/18 A01K67/027 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SOOD R ET AL: "CONSTRUCTION OF A 1-MB RESTRICTION-MAPPED COSMID CONTIG CONTAINING THE CANDIDATE REGION FOR THE FAMILIAL MEDITERRANEAN FEVER LOCUS (MEFV) ON CHROMOSOME 16P13.3" GENOMICS, vol. 42, no. 1, 15 May 1997, pages 83-95, XP002066933 see table 1 ---	6,9
P,X	THE INTERNATIONAL FMF CONSORTIUM: "ANCIENT MISSENSE MUTATIONS IN A NEW MEMBER OF THE RORET GENE FAMILYARE LIKELY TO CAUSE FAMILIAL MEDITERRANEAN FEVER" CELL, vol. 90, no. 4, 22 August 1997, pages 797-807, XP002066936 see page 800 --- -/--	1-4, 6-16, 18-20, 28-30, 32-34, 36-43

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 January 1999	03/02/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Smalt, R

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int. onal Application No
PCT/US 98/17255

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	FRENCH FMF CONSORTIUM: "A CANDIDATE GENE FOR FAMILIAL MEDITERRANEAN FEVER" NATURE GENETICS, vol. 17, no. 1, 1 September 1997, pages 25-31, XP002066935 see the whole document	1-4, 6-16, 18-20, 28-30, 32-34, 36-43
P,X	BERNOT, A. ET AL.: "A transcriptional map of the FMF region." GENOMICS, vol. 50, 1998, pages 147-160, XP002090815 see the whole document -& DATABASE EMBL - R55U031 Entry Hsaj03147, Acc.No. AJ003147, 22 January 1998 BERNOT, A.: "Homo sapiens complete genomic sequence between D16S3070 and D16S3275, containing Familial Mediterranean Fever gene disease" XP002090820 From nt 208600-215910	8
P,X	BERNOT, A. ET AL.: "Non-founder mutations in the MEFV gene establish this gene as the cause of familial mediterranean fever (FMF)" HUMAN MOLECULAR GENETICS, vol. 7, no. 8, August 1998, pages 1317-25, XP002090816 see table 1	12-16, 18-20
T	MCKUSICK, V.A. ET AL.: "Mediterranean fever, familial; MEFV" NCBI - ONLINE MENDELIAN INHERITANCE IN MAN, XP002090817 http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?249100 see the whole document	

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 98/ 17255

Box I Observations wh re certain claims were found uns archable (Continuation f item 1 of first sheet)

This International Search R port ha not been established in respect of certain claims under Articl 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 5,17
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
Claims 5 and 17 have not been searched because none of the claimed sequences are amino acid sequences.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

B x II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees wer accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.